Ferdowsi Civil Engineering, 38, 1, 2025. (1-28)

Ferdowsi University of Mashhad

Ferdowsi Civil Engineering

https://civil-ferdowsi.um.ac.ir

Iranian Concrete Institute

Pavement Engineering Association of Iran

Developing a Robust Data-Driven Model Based on Ground and Satellite Measured Data for Agricultural Drought Prediction in Iran

Research Article

Mahan Azizi[1], Ali Abbasi[2]

DOI: 10.22067/jfcei.2024.88855.1311

1. Introduction

In recent decades, global warming has intensified climate change, resulting in more frequent and severe climatic events, including droughts. The detrimental effects of these phenomena on ecosystems and human societies have become increasingly apparent. Drought, a slow-onset yet destructive natural hazard, is typically defined as a prolonged period of abnormally low precipitation relative to historical averages. Since drought is an inherent climatic phenomenon, effective mitigation of its adverse impacts depends on the availability of timely and accurate forecasts. Consequently, drought indices—which quantify complex interactions between climatic and environmental variables—serve as vital tools for drought monitoring and prediction.

Recent advances in machine learning have significantly enhanced the modeling of drought and climate indices. Integrating machine learning techniques with traditional drought indices offers a robust approach to improving drought forecasting and management. This study aims to develop a high-precision machine learning model for predicting agricultural drought in Iran using the Standardized Precipitation Evapotranspiration Index (SPEI), a widely recognized metric for drought assessment. The proposed model seeks to provide actionable insights to support agricultural planning and drought risk mitigation.

2. Methodology

Data Collection

Meteorological and remote sensing data were collected from 70 stations across Iran, including variables such as air temperature, precipitation, land surface temperature, and vegetation indices. Ground-based measurements were supplemented with satellite-derived datasets to ensure comprehensive spatial coverage.

Feature Selection

To optimize model performance, two feature selection techniques were employed:

1. **Filter Method**: Identified predictors with the strongest statistical correlations to the target SPEI index for use as model inputs.
2. **LASSO Regression**: Reduced model complexity by applying regularization constraints, thereby eliminating less influential variables.

Machine Learning Algorithms

Four algorithms were implemented for SPEI prediction:

1. **Bias-Corrected Random Forest**: A modified Random Forest variant that incorporates bias correction to improve accuracy under diverse climatic conditions.
2. **Random Forest**: An ensemble method aggregating predictions from multiple decision trees to enhance stability and reduce overfitting.
3. **Support Vector Machine (SVM)**: A kernel-based algorithm that classifies drought/non-drought events by mapping data into higher-dimensional feature spaces.
4. **Multilayer Perceptron (MLP)**: A feedforward neural network capable of learning non-linear patterns through iterative training on hidden layers.

3. Model Evaluation

Model performance was assessed using four metrics:

- **Mean Squared Error (MSE)** and **Root MSE (RMSE)**: Quantified prediction errors.
- **Mean Absolute Error (MAE)**: Measured average deviation from observed values.
- **R-squared (R^2)**: Evaluated the proportion of variance explained by the model.

4. Results

As demonstrated in Tables 1 and 2, the Bias-Corrected

[*] Manuscript received July 8, 2024, Revised August 18, 2024, Accepted November 12, 2024.

[1] M.Sc. graduate of Water Resources Engineering and Management, Department of Civil Engineering, Faculty of Engineering, Ferdowsi University of Mashhad, Mashhad, Iran.

[2] Corresponding Author: Assistant Professor, Department of Civil Engineering, Faculty of Engineering, Ferdowsi University of Mashhad, Mashhad, Iran. **Email**: aabbasi@um.ac.ir

Random Forest outperformed other models in SPEI prediction:

- **Cluster 1 (Arid regions)**: Achieved $R^2 = 0.9858$ and RMSE = 0.1190, reflecting exceptional accuracy.
- **Cluster 2 (Semi-arid/humid regions)**: Yielded $R^2 = 0.9809$ and RMSE = 0.1375, indicating strong predictive capability across diverse climates.

Table 1: Results of Drought Prediction Modeling in Cluster 1 Using Machine Learning Algorithms

Machine Learning Algorithm	Feature Selection Method	R^2	RMSE	MSE	MAE
Multilayer Perceptron	Filter	0.9834	0.1286	0.0165	0.0894
	LASSO	0.9852	0.1210	0.0147	0.0853
Support Vector Machine	Filter	0.9822	0.1330	0.0177	0.0915
	LASSO	0.9848	0.1229	0.0151	0.0839
Random Forest	Filter	0.9832	0.1302	0.0170	0.0911
	LASSO	0.9858	0.1191	0.0142	0.0820
Bias-Corrected Random Forest	Filter	0.9841	0.1263	0.0160	0.0879
	LASSO	0.9858	0.1190	0.0142	0.0821

Table 2: Results of Drought Prediction Modeling in Cluster 2 Using Machine Learning Algorithms

Machine Learning Algorithm	Feature Selection Method	R^2	RMSE	MSE	MAE
Multilayer Perceptron	Filter	0.9770	0.1512	0.0229	0.1141
	LASSO	0.9876	0.1457	0.0213	0.1088
Support Vector Machine	Filter	0.9763	0.1534	0.0235	0.1144
	LASSO	0.9803	0.1401	0.0197	0.1045
Random Forest	Filter	0.9761	0.1578	0.0250	0.1164
	LASSO	0.9800	0.1404	0.0198	0.1059
Bias-Corrected Random Forest	Filter	0.9779	0.1496	0.0225	0.1109
	LASSO	0.9809	0.1375	0.0189	0.1031

The bias-corrected Random Forest model demonstrated superior performance compared to other algorithms, owing to its capacity to effectively integrate diverse datasets while applying systematic bias corrections. Subsequently, the optimal model's outputs were utilized to generate agricultural drought zoning maps based on the SPEI index. These maps serve as valuable tools for identifying drought-prone regions and evaluating drought severity across different areas.

The resulting agricultural drought zoning maps, produced from the model's seasonal predictions, provide detailed insights into the spatial distribution and intensity of drought conditions throughout Iran. As illustrated in Figure 1, the analysis reveals distinct seasonal patterns: northern regions (Gilan and Mazandaran) exhibit relatively wet conditions during spring, while central and southern areas experience more severe drought. Furthermore, Figure 2 highlights how rising temperatures and reduced precipitation in summer exacerbate drought severity, particularly in parts of Mazandaran, Gilan, Ardabil, and East Azerbaijan.

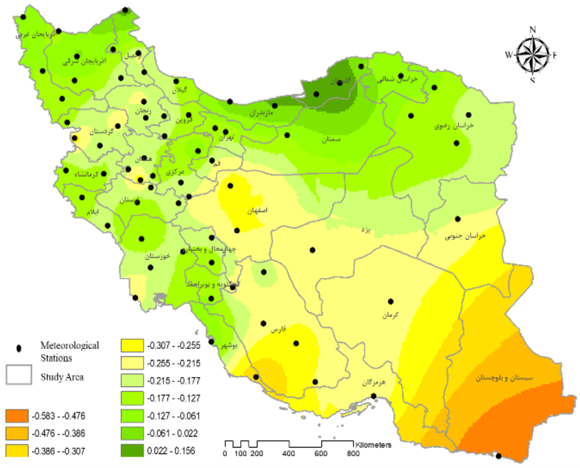


Figure 1: Agricultural Drought Zoning Map in Spring 2021

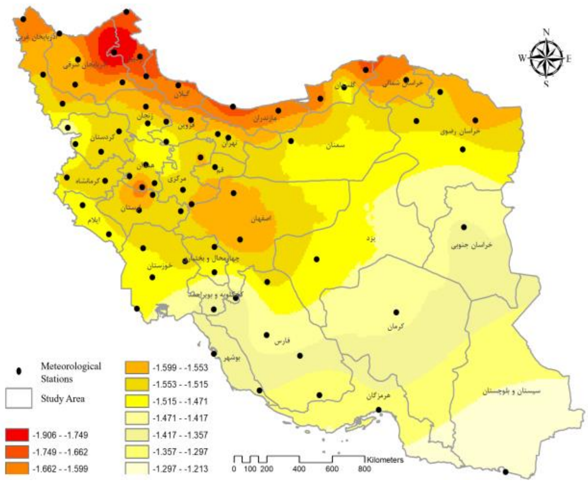


Figure 2: Agricultural Drought Zoning Map in Summer 2021

5- Conclusion

This study demonstrates that integrating multi-source data with the Bias-Corrected Random Forest algorithm significantly enhances drought prediction accuracy. The resulting drought zoning maps, derived from the optimized model, provide policymakers and agricultural experts with actionable insights to improve water resource management and strategic planning.

The findings indicate that this modeling approach can generate precise drought forecasts for diverse regions across Iran. By enabling more effective resource allocation and drought mitigation strategies, the model has the potential to reduce adverse impacts on agricultural productivity and the national economy.

مهندسی عمران فردوسی،۳۸، ۱، ۱۴۰۴. (۲۸-۱)

مهندسی عمران فردوسی

https://civil-ferdowsi.um.ac.ir/

انجمن بتن ایران

انجمن مهندسی روسازی ایران

توسعه یک روش کارآمد مبتنی بر آموزش ماشین و داده‌های زمینی و ماهواره‌ای برای پیش‌بینی و پهنه‌بندی خشکسالی کشاورزی در ایران*

مقاله پژوهشی

ماهان عزیزی[۱]　　　علی عباسی[۲]

DOI: 10.22067/jfcei.2024.88855.1311

چکیده خشکسالی یکی از خطرناک‌ترین وقایع طبیعی برای هر کشور به حساب می‌آید. به همین علت نظارت و پیش‌بینی این پدیده در دنیای امروز از اهمیت بسیار بالایی برخوردار است. در این پژوهش، از شاخص خشکسالی *SPEI* که از معروف‌ترین شاخص‌های خشکسالی به شمار می‌رود، برای پیش‌بینی خشکسالی استفاده شده است. برای محاسبه این شاخص، ترکیبی از داده‌های زمینی و سنجش از دور مورد استفاده قرار گرفته است. با توجه به وزن مختلف هر کدام از این داده‌ها در محاسبه شاخص خشکسالی، ابتدا پارامترهای مؤثر توسط روش‌های انتخاب پارامتر مؤثر مانند روش فیلتر و روش لاسو برگزیده شده و به عنوان پارامترهای ورودی مدل در نظر گرفته شدند. سپس با استفاده از چهار روش یادگیری ماشین شامل جنگل تصادفی تصحیح بایاس شده، جنگل تصادفی، ماشین بردار پشتیبان و پرسپترون چندلایه شاخص خشکسالی محاسبه شده است. برای اعتبارسنجی نتایج هر کدام از این مدل‌ها، از شاخص‌های *RMSE*، R^2، *MSE* و *MAE* استفاده شده و سپس بر اساس مقادیر این شاخص‌ها، مدل جنگل تصادفی تصحیح بایاس شده با مقادیر $R^2 = 0.9858$ و $RMSE = 0.1190$ برای خوشه ۱ و $R^2 = 0.9809$ و $RMSE = 0.1375$ برای خوشه ۲ به عنوان مدل دارای بهترین عملکرد انتخاب شده است. در نهایت از نتایج مدل بهینه برای تهیه نقشه‌های پهنه‌بندی خشکسالی به منظور شناسایی و طبقه‌بندی مناطق مستعد به شرایط خشکسالی استفاده گردید. این نقشه‌ها می‌توانند اطلاعات ارزشمندی را در رابطه با توزیع و شدت خشکسالی در مناطق مختلف در اختیار ذی‌نفعان و متخصصان قرار دهند.

واژه‌های کلیدی خشکسالی، سنجش از دور، یادگیری ماشین، شاخص خشکسالی SPEI، نقشه پهنه‌بندی خشکسالی.

Developing a Robust Data-Driven Model Based on Ground and Satellite Measured Data for Agricultural Drought Prediction in Iran

Mahan Azizi　　　Ali Abbasi

Abstract *Drought is considered one of the most hazardous natural phenomena for any country. Therefore, monitoring and forecasting this phenomenon are of critical importance in today's world. To this end, the Standardized Precipitation Evapotranspiration Index (SPEI), one of the most applied drought indices, was employed for drought prediction. To calculate this index, a combination of ground-based and remote sensing data was utilized. Given the different weights of each of these data types in calculating the drought index, the most relevant parameters were first selected using feature selection methods such as the Filter method and the LASSO method and were considered as input parameters for the model. Furthermore, artificial intelligence was utilized to apply various machine learning algorithms, resulting in the development of several models. These algorithms included Bias-Corrected Random Forest, Random Forest, Support Vector Machine, and Multilayer Perceptron. To validate the results of each of these models, indices such as RMSE, R^2, MSE, and MAE were used. Based on the values of these indices, the Bias-Corrected Random Forest model with $R^2 = 0.9858$ and RMSE = 0.1190 for cluster 1 and $R^2 = 0.9809$ and RMSE = 0.1375 for cluster 2 was selected as the best-performing model. Finally, the results of the optimized model were used to create drought zoning maps to identify and classify areas prone to drought conditions. These maps can provide valuable information on the distribution and intensity of drought across different regions to stakeholders and experts.*

Key Words Drought, Remote sensing, Machine learning, SPEI drought index, Drought maps

* تاریخ دریافت مقاله ۱۴۰۳/۴/۱۸ و تاریخ پذیرش آن ۱۴۰۳/۸/۲۲ می‌باشد.

(۱) کارشناس ارشد گرایش مهندسی و مدیریت منابع آب، گروه عمران، دانشکده مهندسی، دانشگاه فردوسی مشهد، مشهد، ایران.

(۲) نویسنده مسئول: استادیار گروه عمران، دانشکده مهندسی، دانشگاه فردوسی مشهد، مشهد، ایران.

Email: aabbasi@um.ac.ir

مقدمه

امروزه تغییر اقلیم یکی از معضلات بزرگ جهان به شمار می‌رود. در دهه‌های اخیر پیرو گرمایش زمین و افزایش میانگین دمای هوا، فجایع اقلیمی نظیر خشکسالی نیز به گونه‌ای شدت پیدا کرده‌اند که اثرات آن‌ها به وضوح بر اکوسیستم و زندگی انسان قابل مشاهده است. خشکسالی پدیده‌ای خزنده و زیان‌بار است که عموما به حالتی اطلاق می‌شود که میزان بارندگی در یک مدت زمان مشخص از مقدار آستانه مشخص شده کمتر باشد. از آن‌جا که خشکسالی یک پدیده طبیعی است، مقابله با اثرات منفی آن تنها در صورتی امکان‌پذیر است که در مورد وقوع آن هشدار قبلی و به موقعی داده شود [1]. به همین جهت وجود یک سیستم کارآمد برای پیش‌بینی این پدیده امری ضروری و اجتناب‌ناپذیر است. بدین منظور استفاده از شاخص‌های خشکسالی که رابطه پیچیده‌ای با اقلیم و محیط دارند، به عنوان روشی مؤثر برای پیش‌بینی و پایش خشکسالی شناخته می‌شود [2]. به طور کلی شاخص‌های خشکسالی ابزارهایی هستند که در تشخیص شروع، شدت و پایان خشکسالی به کمک انسان می‌آیند. نحوه عملکرد این شاخص‌ها بدین صورت می‌باشد که هزاران بیت داده در مورد بارندگی، برف، پوشش گیاهی و... را دریافت کرده، آن‌ها را در بازه‌های زمانی مختلف تجزیه و تحلیل کرده و در نهایت به یک تصویر قابل درک تبدیل می‌کنند.

یکی از مهم‌ترین مراحل در نظارت و تحلیل خشکسالی، استفاده از شاخص‌های خشکسالی مناسب است. به طور کلی، شاخص خشکسالی یک متغیر اصلی برای تعیین مدت و شدت خشکسالی و ارزیابی تأثیرات آن است. با این حال، به دلیل پیچیدگی تعاریف خشکسالی، ارزیابی جامع و سیستماتیک ویژگی‌های خشکسالی در یک منطقه مشخص تنها با یک شاخص امکان‌پذیر نیست. در همین راستا، صدها شاخص خشکسالی توسط محققان در سراسر جهان توسعه یافته است. این شاخص‌ها را می‌توان به دو دسته تقسیم کرد: شاخص‌های مبتنی بر زمین، و شاخص‌های سنجش از دوری.

شاخص‌های خشکسالی مبتنی بر زمین معمولا از اندازه‌گیری‌های زمینی متغیرهای هواشناسی، مانند بارندگی و دما، به دست می‌آیند. این شاخص‌ها قادرند به طور دقیق و مؤثر خشکسالی را در اطراف ایستگاه‌های اقلیمی نظارت کنند. برخی از آن‌ها شامل شاخص PDSI[3]، شاخص SPI[4] و شاخص SPEI[5] هستند.

اما برای پوشش گستره مکانی شاخص‌های خشکسالی مبتنی بر زمین، شاخص‌های سنجش از دور بر اساس داده‌های ماهواره‌ای توسعه یافته‌اند. داده‌های سنجش از دور ماهواره‌ای شامل اطلاعاتی از قبیل بارندگی، دما، تبخیر و تعرق، و پوشش گیاهی هستند که همگی پوشش پیوسته و لحظه‌ای دارند. در دهه‌های اخیر، شاخص‌های خشکسالی سنجش از دور متعددی مانند شاخص NDVI [6]، شاخص NDDI [7]، و شاخص NMDI [8] توسعه یافته‌اند. با این حال، با وجود اینکه این شاخص‌ها کمبودهای شاخص‌های خشکسالی مبتنی بر زمین در پوشش مکانی را برطرف می‌کنند، اما از آن‌جا که داده‌های آن‌ها نسبتا کوتاه‌مدت هستند، نمی‌توانند به طور کامل جایگزین مشاهدات مبتنی بر زمین در بحث نظارت بر خشکسالی شوند. علاوه بر این، کیفیت شاخص‌های خشکسالی سنجش از دور می‌تواند تحت تأثیر شرایط جوی (مانند ابرها) نیز قرار گیرد.

به طور کلی، در موضوع نظارت بر خشکسالی، شاخص‌های خشکسالی مبتنی بر زمین معمولا دقیق‌تر از شاخص‌های خشکسالی سنجش از دور در نظر گرفته می‌شوند. این امر عمدتا به این دلیل است که متغیرهای اقلیمی استخراج‌شده از اندازه‌گیری‌های زمینی معمولا دقیق‌تر هستند. علاوه بر این، اکثر شاخص‌های خشکسالی مبتنی بر زمین به طور مستقیم تعادل میان آب و خاک را نظارت می‌کنند نه اطلاعات غیرمستقیم پوشش گیاهی. همچنین، بسیاری از شاخص‌های خشکسالی سنجش از دور که به تازگی توسعه یافته‌اند نیز معمولا نیاز به صحت‌سنجی با استفاده از شاخص‌های مبتنی بر زمین دارند تا اعتبار آن‌ها اثبات گردد [9].

در سالیان اخیر کاربردهای مدل‌های یادگیری ماشین، پیشرفت چشمگیری در مدل‌سازی شاخص‌های خشکسالی و اقلیم‌شناسی از خود نشان داده‌اند [10]. لذا به‌کارگیری همزمان مدل‌های یادگیری ماشین و شاخص‌های خشکسالی می‌تواند

ابزار قدرتمندی برای نظارت، پیش‌بینی و مدیریت خشکسالی‌ها فراهم آورد. در همین راستا تلاش‌های بسیاری توسط محققان متعدد صورت پذیرفته است تا بتوان داده‌ها را از منابع مختلف ترکیب کرد و شاخص‌های خشکسالی مبتنی بر زمین را با استفاده از مدل‌های داده‌محور مبتنی بر یادگیری ماشین مانند شبکه‌های عصبی مصنوعی، ماشین بردار پشتیبان، جنگل تصادفی و... بازتولید کرد. به عنوان مثال کوثری و همکاران (۲۰۱۷) [11] در مطالعه‌ای یک روش پیش‌بینی خشکسالی منطقه‌ای بر مبنای شبکه عصبی مصنوعی و شاخص SPI در سری‌های ۳، ۶، ۹، ۱۲، ۱۸ و ۲۴ ماهه در استان فارس ایران برای معرفی مدلی جهت پیش‌بینی خشکسالی ارائه نمودند. در این پژوهش از داده‌های بارندگی ۴۱ ایستگاه باران‌سنجی برای محاسبه مقادیر شاخص بارش استاندارد استفاده شد و آزمایش‌های مکرر و اعتبارسنجی آن‌ها برای به دست آوردن بهترین مدل شبکه عصبی مصنوعی در نظر گرفته شد. نتایج این پژوهش نشان داد مقادیر پیش‌بینی شده شش ماهه اول به طور میانگین ۷۳ درصد تطابق با مقادیر مشاهده شده را داراست. مختارزاد و همکاران (۲۰۱۷) [12] نیز در پژوهشی برای پیش‌بینی خشکسالی از شبکه عصبی مصنوعی، سیستم استنتاج عصبی فازی تطبیقی و ماشین بردار پشتیبان استفاده کرده و نتایج این روش‌ها را با یکدیگر مقایسه کردند. آن‌ها بدین منظور از تمام داده‌های هواشناسی (دما، رطوبت و بارندگی) ایستگاه بجنورد واقع در ایران، میان سال‌های ۱۹۸۴ تا ۲۰۱۲ در مقیاس زمانی سه ماهه استفاده گردید تا در نهایت شاخص SPI به عنوان خروجی به دست آید. نتایج نشان داد که الگوریتم ماشین بردار پشتیبان مدل دقیق‌تری را برای پیش‌بینی ارائه می‌دهد. در پژوهشی دیگر عباسی و همکاران (۲۰۲۰) [13] به بررسی وضعیت خشکسالی در ایستگاه سینوپتیک سقز در مقیاس‌های زمانی مختلف با استفاده از شاخص بارش ـ تبخیر و تعرق استاندارد شده و الگوریتم ماشین بردار پشتیبان با سه تابع هسته‌ای خطی، چند جمله‌ای و پایه شعاعی و شبکه بیزین پرداختند. برای این منظور از شاخص SPEI در مقیاس‌های زمانی کوتاه‌مدت یک و سه ماهه، میان‌مدت شش و ۱۲ ماهه و بلندمدت ۲۴ و ۴۸ ماهه در طی دوره آماری ۴۹ ساله برای پایش وضعیت خشکسالی در این ایستگاه استفاده شد. نتایج نشان داد که تابع هسته‌ای خطی در روش ماشین بردار پشتیبان نسبت به دو تابع دیگر دقت بیشتری داشته است.

توفانر و اوزبیاز (۲۰۲۰) [14] نیز در پژوهشی به تخمین شاخص شدت خشکسالی پالمر در منطقه آدیامان واقع در کشور ترکیه پرداختند. در این پژوهش پارامترهای شاخص شدت خشکسالی پالمر به کمک چهار روش رگرسیونی و یادگیری ماشین شامل رگرسیون خطی، شبکه عصبی مصنوعی، ماشین بردار پشتیبان و درخت تصمیم تخمین زده شدند. نتایج پژوهش نشان داد که الگوریتم‌های استفاده شده با شاخص شدت خشکسالی پالمر سازگاری داشته و مدل‌های رگرسیونی محاسبات تخمین شاخص شدت خشکسالی پالمر را تسهیل می‌کنند که این امر برای مطالعات آنالیز و پیش‌بینی خشکسالی بسیار کارآمد می‌باشد. فنگ و همکاران (۲۰۱۹)[15] نیز در مطالعه‌ای از سی عامل خشکسالی مبتنی بر روش‌های سنجش از دور برای بازتولید شاخص SPEI میان سال‌های ۲۰۰۱ تا ۲۰۱۷ برای منطقه نیو ساوث ولز در جنوب شرقی استرالیا استفاده کردند. در این پژوهش سه روش پیشرفته یادگیری ماشین شامل جنگل تصادفی تصحیح بایاس شده، ماشین بردار پشتیبان و شبکه عصبی پرسپترون چند لایه، به عنوان مدل‌های رگرسیون در این روش به کار گرفته شدند. نتایج نشان داد که مدل جنگل تصادفی تصحیح بایاس شده از دو مدل دیگر برای پیش‌بینی شاخص عملکرد بهتری دارد.

موضوع خشکسالی و سازگاری با آن در ایران نیز از اهمیت بالایی برخوردار است. مجموعه داده‌های شاخص PDSI از سال ۱۹۵۱ تا ۲۰۰۵ نشان داده است که شدت خشکسالی در ایران، به ویژه در شمال غرب و شمال شرق کشور افزایش یافته است [16]. به گونه‌ای که تنها در سال ۲۰۰۱، حدود هشت میلیون هکتار از اراضی کشاورزی ایران تحت تأثیر خشکسالی قرار گرفت و میلیون‌ها دلار خسارت به کشور وارد کرد [17]. لذا لزوم ایجاد مدلی کارآمد برای پیش‌بینی این مخاطره طبیعی در ایران، امری اجتناب‌ناپذیر به نظر می‌آید. بنابراین، می‌توان اهداف اصلی این پژوهش را بدین ترتیب بیان کرد:

۱. تهیه نقشه پهنه‌بندی خشکسالی کشاورزی.
۲. یافتن مناسب‌ترین پارامترها و شاخص‌های خشکسالی مؤثر در ایجاد مدل پیش‌بینی خشکسالی.
۳. اعمال الگوریتم‌های مختلف یادگیری ماشین و یافتن دقیق‌ترین و بهینه‌ترین الگوریتم.

ضمنا می‌توان به طور کلی مهم‌ترین نوآوری‌های این پژوهش را که حد تمایزی بین این مطالعه و سایر مطالعات پیشین انجام شده بر روی منطقه مورد مطالعه می‌باشند، در این موارد خلاصه کرد:

۱. بهره گرفتن همزمان از داده‌های زمینی و داده‌های سنجش از دوری به منظور افزایش دقت مدل‌سازی با توجه گستردگی منطقه مورد مطالعه.
۲. افزایش دقت مدل‌سازی در مناطق با اهمیت کشاورزی بالاتر از طریق تخصیص ایستگاه‌های جمع‌آوری داده بیشتر به استان‌های با سهم بالاتر از مجموع مساحت اراضی زراعی کشور.
۳. بهره گرفتن از دو روش متفاوت برای انتخاب پارامترهای مؤثر و چهار الگوریتم یادگیری ماشین متفاوت به منظور انتخاب دقیق‌ترین مدل ایجاد شده.

مواد و روش‌ها

منطقه مورد مطالعه

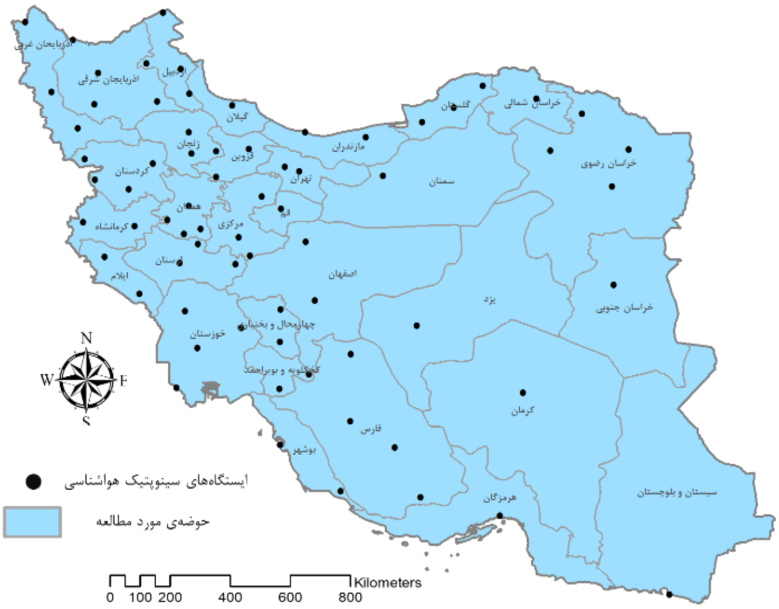
منطقه مورد مطالعه انتخابی در این پژوهش کشور ایران می‌باشد. ایران با مساحتی بیش از ۱/۶ میلیون کیلومتر مربع، کشوری پهناور محسوب می‌شود. به همین علت و نیز به دلیل گستردگی این مساحت بین عرض جغرافیایی ۲۵ الی ۴۰ درجه شمالی، مناطق مختلف کشور شرایط اقلیمی و آب و هوایی متنوعی را در فصول مختلف تجربه می‌کنند. به همین منظور، اهمیت کشاورزی و در پی آن اهمیت اعمال دقت دوچندان در پیش‌بینی خشکسالی در مناطق مختلف کشور، با توجه به تفاوت‌های چشمگیر میان این مناطق، بسیار حائز اهمیت است.

لذا در ابتدا با توجه به تقسیمات رسمی استانی در کشور ایران، که به تعداد ۳۱ استان تفکیک شده است، با در نظر گرفتن معیار سهم هر استان از کل مساحت اراضی زراعی کشور، یک یا چند ایستگاه هواشناسی برای هر استان معین و انتخاب شده و سپس تمامی محاسبات برای محل قرارگیری این ایستگاه‌ها صورت پذیرفت. بر اساس اطلاعات به دست آمده از نشریه سهم و رتبه استان‌ها بر اساس متغیرهای مهم بخش کشاورزی [18] که توسط مرکز آمار ایران منتشر گردیده است، می‌توان استان‌های کشور را بر اساس درصد مساحت اراضی کشاورزی استان از مساحت کل اراضی کشاورزی کشور، بین چهار دسته مختلف تقسیم‌بندی کرد. در ادامه، پس از تخصیص دادن تمامی استان‌ها به یکی از چهار دسته تعریف شده، از هر استان بنا به تعلق به هر یک از دسته‌های ذکر شده به ترتیب ۴، ۳، ۲ و ۱ ایستگاه هواشناسی انتخاب گردید. انتخاب این ایستگاه‌ها از میان مجموع ۴۵۵ ایستگاه سینوپتیک متعلق به سازمان هواشناسی کشور صورت گرفته است. جدول (۱) نمایانگر جزئیات این دسته‌بندی می‌باشد. **Error! Reference source not found.** نیز محل‌های قرارگیری ایستگاه‌های منتخب را در سراسر کشور نشان می‌دهد. لازم به ذکر است که از ۷۵ ایستگاه پیش‌بینی شده برای این پژوهش، مجموعا چهار ایستگاه از استان‌های زنجان، خراسان شمالی، همدان و کرمان به علت عدم پوشش‌دهی کامل بازه زمانی مورد نظر در زمینه فراهم بودن داده از سوی سازمان هواشناسی کشور و نیز یک ایستگاه از استان سیستان و بلوچستان به علت عدم پوشش‌دهی منطقه مورد نظر از سوی سنجنده مادیس، مورد تأمین داده قرار نگرفته و به ناچار حذف گردیدند. مشخصات و اطلاعات کلی این ایستگاه‌ها در ادامه و در قالب جدول (۲) ارائه شده است.

در این پژوهش، بازه زمانی ۲۲ ساله از دی ماه ۱۳۷۹ هجری شمسی مصادف با ژانویه ۲۰۰۱ میلادی لغایت دی ماه ۱۴۰۱ شمسی مصادف با ژانویه ۲۰۲۳ میلادی انتخاب شده و مورد مطالعه قرار گرفته است.

جدول ۱ جزئیات دسته‌بندی استان‌های کشور بر اساس اهمیت کشاورزی

نام دسته	استان	سهم ایستگاه‌های استان از مجموع ایستگاه‌های مورد مطالعه (درصد)
الف (سهم بیش از ۵ درصد از کل اراضی زراعی کشور)	خراسان رضوی	۵/۷۱٪
	خوزستان	۵/۷۱٪
	آذربایجان شرقی	۵/۷۱٪
	کردستان	۵/۷۱٪
	فارس	۵/۷۱٪
	اردبیل	۵/۷۱٪
	همدان	۴/۲۹٪
	زنجان	۴/۲۹٪
ب (سهم بین ۳ تا ۵ درصد از کل اراضی زراعی کشور)	لرستان	۴/۲۹٪
	آذربایجان غربی	۴/۲۹٪
	کرمانشاه	۴/۲۹٪
	مرکزی	۴/۲۹٪
	گلستان	۴/۲۹٪
پ (سهم بین ۱ تا ۳ درصد از کل اراضی زراعی کشور)	قزوین	۲/۸۶٪
	اصفهان	۲/۸۶٪
	بوشهر	۲/۸۶٪
	ایلام	۲/۸۶٪
	مازندران	۲/۸۶٪
	کهگیلویه و بویراحمد	۲/۸۶٪
	چهارمحال و بختیاری	۲/۸۶٪
	خراسان شمالی	۱/۴۳٪
	کرمان	۱/۴۳٪
ت (سهم کمتر از ۱ درصد از کل اراضی زراعی کشور)	سیستان و بلوچستان	۱/۴۳٪
	گیلان	۱/۴۳٪
	خراسان جنوبی	۱/۴۳٪
	سمنان	۱/۴۳٪
	تهران	۱/۴۳٪
	هرمزگان	۱/۴۳٪
	البرز	۱/۴۳٪
	یزد	۱/۴۳٪
	قم	۱/۴۳٪



شکل ۱ منطقه مورد مطالعه به همراه محل قرارگیری ایستگاه‌های هواشناسی مورد استفاده

جدول ۲ مشخصات کلی ایستگاه‌های سینوپتیک مورد استفاده در این پژوهش

استان	نام ایستگاه	طول جغرافیایی	عرض جغرافیایی	میانگین بارندگی سالانه (میلی‌متر)	میانگین دمای سالانه (درجه سانتی‌گراد)
خراسان رضوی	مشهد	۵۹/۶	۳۶/۲	۲۲۷	۱۵/۹
	سبزوار	۵۷/۶	۳۶/۲	۱۷۶	۱۸/۲
	قوچان	۵۸/۴	۳۷/۱	۳۲۴	۱۲/۵
	تربت حیدریه	۵۹/۲	۳۵/۳	۲۲۱	۱۴/۷
خوزستان	اهواز	۴۸/۷	۳۱/۳	۱۹۳	۲۶/۷
	آبادان	۴۸/۲	۳۰/۴	۱۳۴	۲۶/۶
	ایذه	۴۹/۹	۳۱/۸	۵۷۰	۲۳/۸
	صفی‌آباد (دزفول)	۴۸/۴	۳۲/۳	۲۸۶	۲۴/۴
آذربایجان شرقی	تبریز	۴۶/۲	۳۸/۱	۲۵۷	۱۳/۴
	جلفا	۴۵/۶	۳۸/۹	۲۲۴	۱۵/۵
	میانه	۴۷/۷	۳۷/۴	۲۶۲	۱۴/۴
	مراغه	۴۶/۱	۳۷/۳	۲۵۲	۱۳/۷
کردستان	سنندج	۴۷/۰	۳۵/۳	۳۵۰	۱۴/۶
	بانه	۴۵/۹	۳۶/۰	۶۹۰	۱۴/۳
	بیجار	۴۷/۶	۳۵/۹	۳۲۰	۱۲/۰
	مریوان	۴۶/۲	۳۵/۵	۸۶۱	۱۳/۶
فارس	شیراز	۵۲/۶	۲۹/۵	۲۸۴	۱۹/۱
	لار	۵۴/۴	۲۷/۷	۱۸۸	۲۴/۸
	آباده	۵۲/۶	۳۱/۲	۱۳۷	۱۴/۹
	فسا	۵۳/۷	۲۸/۹	۲۵۶	۱۹/۹

ادامه جدول ۳ مشخصات کلی ایستگاه‌های سینوپتیک مورد استفاده در این پژوهش

استان	نام ایستگاه	طول جغرافیایی	عرض جغرافیایی	میانگین بارندگی سالانه (میلی‌متر)	میانگین دمای سالانه (درجه سانتی‌گراد)
اردبیل	اردبیل	۴۸/۳	۳۸/۲	۲۵۹	۱۰/۱
	مشکین‌شهر	۴۷/۵	۳۸/۴	۳۷۶	۱۱/۲
	خلخال	۴۸/۵	۳۷/۶	۳۷۱	۹/۲
	پارس‌آباد	۴۷/۹	۳۹/۶	۲۷۱	۱۵/۱
همدان	فرودگاه همدان	۴۸/۵	۳۴/۹	۳۰۲	۱۲/۴
	نهاوند	۴۸/۴	۳۴/۲	۳۶۵	۱۴/۹
	ملایر	۴۸/۸	۳۴/۳	۳۴۰	۱۳/۹
زنجان	زنجان	۴۸/۵	۳۶/۷	۲۸۶	۱۱/۹
	خرم‌دره	۴۹/۲	۳۶/۲	۲۹۷	۱۲/۷
	خدابنده	۴۸/۶	۳۶/۱	۳۸۴	۱۱/۵
لرستان	خرم‌آباد	۴۸/۳	۳۳/۴	۴۶۱	۱۷/۵
	بروجرد	۴۸/۸	۳۳/۹	۴۶۲	۱۵/۳
	الیگودرز	۴۹/۷	۳۳/۴	۳۹۱	۱۳/۲
آذربایجان غربی	ارومیه	۴۵/۱	۳۷/۷	۳۰۳	۱۱/۹
	مهاباد	۴۵/۷	۳۶/۸	۳۷۸	۱۳/۸
	ماکو	۴۴/۴	۳۹/۴	۳۲۲	۱۱/۲
کرمانشاه	کرمانشاه	۴۷/۲	۳۴/۴	۳۸۹	۱۶/۱
	سرپل‌ذهاب	۴۵/۹	۳۴/۵	۳۸۷	۲۰/۶
	کنگاور	۴۸/۰	۳۴/۵	۳۷۲	۱۳/۶
مرکزی	اراک	۴۹/۸	۳۴/۱	۳۰۱	۱۴/۶
	ساوه	۵۰/۴	۳۵/۱	۱۸۴	۱۹/۹
	خمین	۵۰/۱	۳۳/۶	۲۹۲	۱۵/۶
گلستان	گرگان	۵۴/۴	۳۶/۹	۴۹۰	۱۷/۷
	گنبدکاووس	۵۵/۲	۳۷/۳	۴۳۰	۱۹/۳
	مراوه‌تپه	۵۵/۹	۳۷/۸	۳۶۳	۱۸/۰
قزوین	قزوین	۵۰/۰	۳۶/۲	۳۰۱	۱۴/۴
	آوج	۴۹/۲	۳۵/۶	۳۷۳	۱۱/۹
اصفهان	اصفهان	۵۱/۷	۳۲/۵	۱۳۱	۱۷/۵
	کاشان	۵۱/۵	۳۴/۰	۱۲۳	۱۹/۸
بوشهر	فرودگاه بوشهر	۵۰/۸	۲۹/۰	۲۲۴	۲۵/۷
	جم	۵۲/۴	۲۷/۸	۳۱۳	۲۶/۱
ایلام	ایلام	۴۶/۴	۳۳/۶	۵۰۹	۱۷/۰
	دهلران	۴۷/۳	۳۲/۷	۲۶۱	۲۶/۹
مازندران	ساری	۵۳/۰	۳۶/۵	۷۴۷	۱۷/۸
	نوشهر	۵۱/۵	۳۶/۷	۱۳۱۸	۱۶/۸

ادامه جدول ۴ مشخصات کلی ایستگاه‌های سینوپتیک مورد استفاده در این پژوهش

استان	نام ایستگاه	طول جغرافیایی	عرض جغرافیایی	میانگین بارندگی سالانه (میلی‌متر)	میانگین دمای سالانه (درجه سانتی‌گراد)
کهگیلویه و بویراحمد	یاسوج	۵۱/۶	۳۰/۷	۷۷۱	۱۵/۱
	دوگنبدان	۵۰/۸	۳۰/۳	۴۰۹	۲۳/۴
چهارمحال و بختیاری	شهرکرد	۵۰/۸	۳۲/۳	۳۰۹	۱۱/۸
	لردگان	۵۰/۸	۳۱/۵	۵۴۸	۱۶/۰
خراسان شمالی	بجنورد	۵۷/۳	۳۷/۵	۲۴۳	۱۳/۵
کرمان	کرمان	۵۷/۰	۳۰/۳	۱۱۱	۱۷/۴
سیستان و بلوچستان	چابهار	۶۰/۷	۲۵/۳	۱۲۰	۲۶/۷
گیلان	رشت	۴۹/۶	۳۷/۳	۱۲۸۳	۱۶/۵
خراسان جنوبی	بیرجند	۵۹/۳	۳۲/۹	۱۳۲	۱۷/۰
سمنان	سمنان	۵۳/۴	۳۵/۶	۱۳۰	۱۸/۸
تهران	فرودگاه مهرآباد	۵۱/۳	۳۵/۷	۲۳۵	۱۸/۶
هرمزگان	بندرعباس	۵۶/۴	۲۷/۲	۱۳۱	۲۷/۳
البرز	کرج	۵۰/۹	۳۵/۸	۲۶۱	۱۶/۱
یزد	یزد	۵۴/۳	۳۱/۹	۵۱	۲۰/۷
قم	قم	۵۰/۹	۳۴/۸	۱۴۱	۱۹/۰

داده‌ها

خشکسالی به مانند اغلب رویدادهای اقلیمی پدیده‌ای پیچیده است که تحت تأثیر متغیرها و عوامل متعددی قرار دارد. به همین علت، درک این عوامل و متغیرها برای درک علل، اثرات و راهبردهای کاهش احتمالی اثرات چنین رویدادهایی بسیار حائز اهمیت است. در این میان، خشکسالی به تنهایی وابسته به طیف بسیار گسترده‌ای از این متغیرها و عوامل می‌باشد. لذا واضح است که مورد ارزیابی قرار دادن شرایط خشکسالی در حالی که تمامی عوامل تأثیرگذار بر آن در نظر گرفته شده و در پیش‌بینی آن لحاظ گردیده، امری طولانی، بسیار هزینه‌بر و نیازمند انجام محاسبات بسیار سنگین است. با این حال، در این پژوهش به دلیل گستردگی منطقه مورد مطالعه و نیز به منظور افزایش دقت عملکرد مدل، از متغیرها و شاخص‌های خشکسالی مبتنی بر زمین و سنجش از دوری متعددی استفاده شده است. تمامی این متغیرها و شاخص‌ها برای بازه زمانی ۲۲ ساله مورد مطالعه، جمع‌آوری گردیده و مورد استفاده قرار گرفته‌اند. خلاصه‌ای از این داده‌ها و شاخص‌ها در قالب جدول (۳) ارائه شده است.

در ادامه به منظور دستیابی به مقیاسی واحد، تمامی متغیرها و شاخص‌های استخراج شده به تفکیک ماهانه تغییر مقیاس داده شده و نیز برای در نظر گرفتن پاسخ تأخیری ناشی از اثر تأخیری یک متغیر بر دیگری، علاوه بر مقادیر ماهانه، میانگین متحرک سه ماهه برای تمامی داده‌ها و شاخص‌های موجود در این مطالعه، محاسبه شده و نهایتا از روش Z-score برای استانداردسازی تمامی این داده‌ها استفاده شده است. Z-score یک تکنیک پرکاربرد برای استانداردسازی داده‌ها است که مقادیر را به گونه‌ای تبدیل می‌کند که میانگین آن‌ها صفر و انحراف از معیار آن‌ها یک باشد. این روش به کمک رابطه (۱) قابل تعریف است. در این رابطه x_s نشان‌دهنده مقادیر استاندارد شده، x نشان‌دهنده مقادیر اولیه، μ میانگین تمام داده‌ها و σ انحراف از معیار داده‌ها می‌باشند.

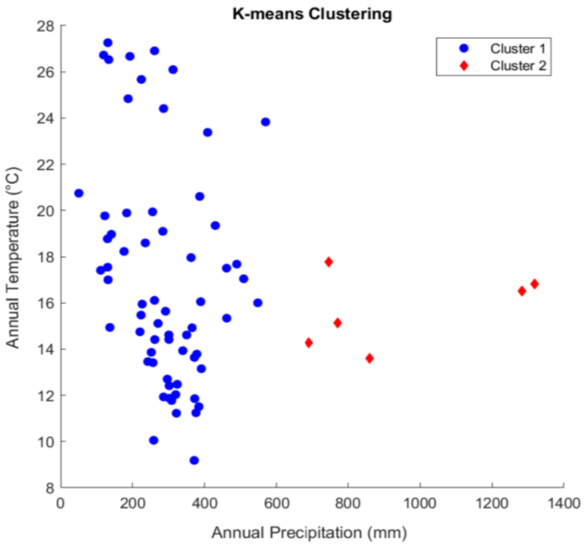
$$x_s = \frac{x - \mu}{\sigma} \quad (۱)$$

جدول ۵ خلاصه داده‌ها و شاخص‌های مورد استفاده

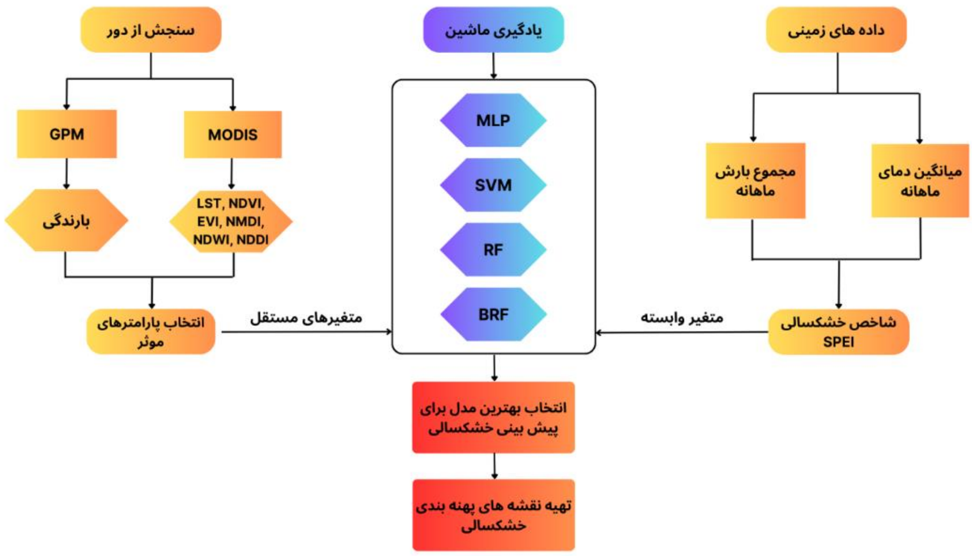
داده‌ها	منبع دسترسی	نام محصول	تفکیک زمانی	تفکیک مکانی
دمای هوا	سازمان هواشناسی کشور [28]	میانگین دمای ماهانه	ماهانه	ایستگاه
بارندگی	سازمان هواشناسی کشور [28]	مجموع بارش ماهانه	ماهانه	ایستگاه
	GPM [29]	GPMv6	نیم ساعته	۱۱۱۳۲ متر
دمای سطح زمین	MODIS [30]	MOD11A1	روزانه	۱۰۰۰ متر
NDVI	MODIS [31]	MOD09GA	روزانه	۵۰۰ متر
EVI	MODIS [31]	MOD09GA	روزانه	۵۰۰ متر
NMDI	MODIS [31]	MOD09GA	روزانه	۵۰۰ متر
NDWI	MODIS [31]	MOD09GA	روزانه	۵۰۰ متر
NDDI	MODIS [31]	MOD09GA	روزانه	۵۰۰ متر

مدل‌سازی

یکی از ملزومات ایجاد یک مدل دقیق و کارآمد در مدل‌سازی، فراهم کردن شرایطی است که پارامترهای دخیل در مدل دارای شرایط و خصوصیات مشابه باشند. خوشه‌بندی تکنیکی برای گروه‌بندی اشیای مشابه یا گروه نقاط می‌باشد که در یادگیری ماشین و تجزیه و تحلیل داده‌ها مورد استفاده قرار می‌گیرد. بنابراین، در اولین گام از فرایند مدل‌سازی، از میان موارد متعدد الگوریتم‌های خوشه‌بندی موجود، الگوریتم K-means [19] که از محبوبیت بالایی نیز برخوردار است، انتخاب شد و مورد استفاده قرار گرفت. در ادامه، پس از بررسی حالات مختلف، تعداد خوشه‌ها دو در نظر گفته شده و سپس از دو معیار میانگین بارش سالانه و میانگین دمای سالانه برای تخصیص هر یک از ایستگاه‌ها به یکی از خوشه‌ها استفاده شد. این الگوریتم شش ایستگاه را به خوشه دوم و ۶۴ ایستگاه را به خوشه اول اختصاص داد. در این خوشه‌بندی، ایستگاه‌های *بانه، مریوان، نوشهر، رشت، ساری و یاسوج* در خوشه دوم و مابقی ایستگاه-ها در خوشه اول قرار گرفتند. همچنین مطابق با این دسته‌بندی، ایستگاه‌های متعلق به خوشه دوم دارای میانگین بارندگی سالانه ۹۴۵ میلی‌متر و میانگین دمای سالانه ۱۵/۷ درجه سانتی‌گراد بودند. این مقادیر برای خوشه اول به ترتیب ۲۸۹ میلی‌متر و ۱۷/۱ درجه سانتی‌گراد بود. این بدین معناست که ایستگاه‌های خوشه دوم به مراتب مرطوب‌تر و دارای دمای نسبتا پایین‌تری نسبت به ایستگاه‌های خوشه اول بودند. از آن‌جایی که غالب مساحت ایران دارای اقلیم نسبتا خشکی می‌باشد، تخصیص بیش از ۹۰ درصد ایستگاه‌های مورد مطالعه به یک خوشه، منطقی به نظر می‌رسد. نتایج این خوشه‌بندی در شکل (۲) قابل مشاهده است. در مرحله بعد، برای هر خوشه، چهار مدل مختلف یادگیری ماشین، یعنی پرسپترون چندلایه (MLP)، ماشین بردار پشتیبان (SVM)، جنگل تصادفی (RF) و جنگل تصادفی تصحیح بایاس شده (BRF) برای پیش‌بینی خشکسالی توسعه داده شدند و بهترین مدل تعیین شد. در نهایت، از نتایج بهترین مدل برای تولید نقشه‌های پهنه‌بندی خشکسالی استفاده گردید. شکل (۳) روند کلی مراحل مختلف این مدل‌سازی را در قالب یک فلوچارت نشان می‌دهد. در ادامه به تفصیل به بررسی روش‌های استفاده شده خواهیم پرداخت.



شکل ۲ نتایج خوشه‌بندی به روش K-means



شکل ۳ فلوچارت مراحل انجام این پژوهش

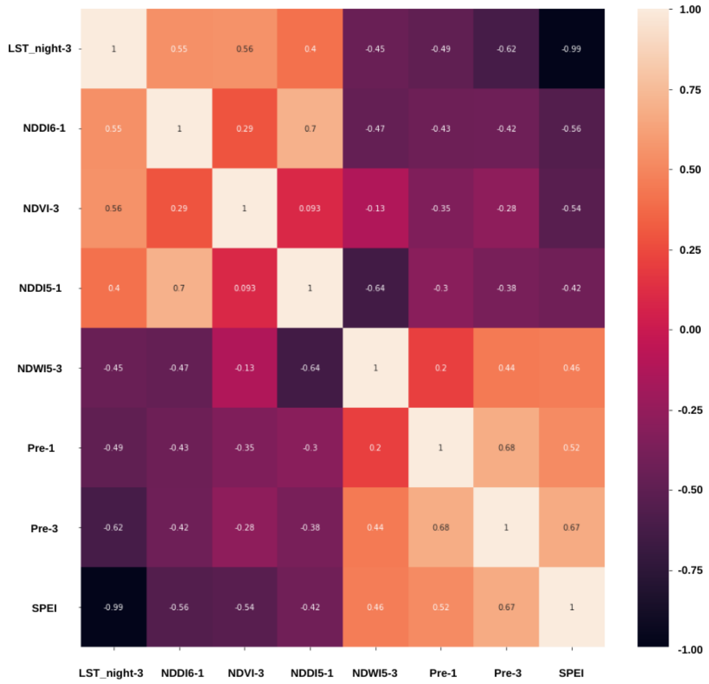
شاخص بارش ـ تبخیر و تعرق استاندارد شده (SPEI)

شاخص SPEI یک شاخص خشکسالی است که اولین بار توسط سرانو و همکاران [5] مورد استفاده قرار گرفت. این شاخص که از تفاضل مقادیر دو پارامتر بارندگی و تبخیر و تعرق پتانسیل برای تعیین شدت و پایش شرایط خشکسالی استفاده می‌کند، یک شاخص استاندارد شده و بدون بعد است که امکان مقایسه شدت خشکسالی در مکان‌ها و دوره‌های زمانی مختلف را فراهم می‌کند. از آن‌جایی که شاخص SPEI از روش Z-score برای استانداردسازی استفاده می‌کند، مقادیر این شاخص عموما بین ۳- تا ۳ قرار می‌گیرد. در این بازه، مقدار صفر بیانگر شرایط عادی است، در حالی که مقادیر منفی و مثبت به ترتیب شرایط خشک و مرطوب را نشان می‌دهند و به طور کلی، مقادیر کوچک‌تر از ۱- به عنوان شرایط خشکسالی در نظر گرفته می‌شود [15]. از آن‌جایی که در پژوهش حاضر، شاخص SPEI به عنوان پارامتر مورد پیش‌بینی (متغیر وابسته) مدل یادگیری ماشین عمل می‌کند، محاسبه این پارامتر با استفاده از داده‌های ایستگاه‌های زمینی، که از دقت و قابلیت اعتماد بالاتری برخوردارند، صورت پذیرفته است. همان‌طور که پیش‌تر نیز اشاره شد، این شاخص از تفاضل مقادیر دو پارامتر بارندگی و تبخیر و تعرق پتانسیل حاصل می‌شود. در این پژوهش از روش تورنث‌ویت [20] برای محاسبه تبخیر و تعرق پتانسیل استفاده شده است. این روش تبخیر و تعرق پتانسیل را بر اساس داده‌های میانگین ماهانه دما و با در نظر گرفتن رابطه بین دما و تبخیر و تعرق محاسبه می‌کند.

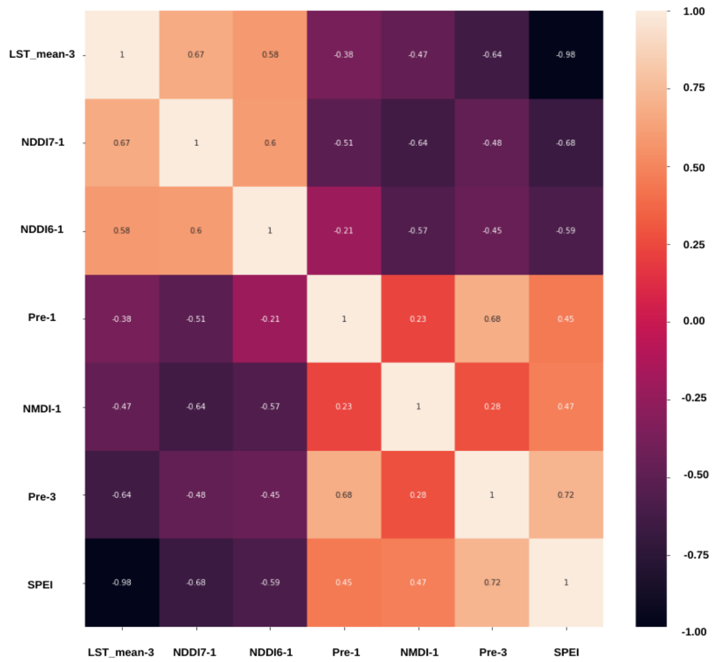
انتخاب پارامترهای مؤثر

انتخاب پارامترهای مؤثر یک گام مهم در ساخت یک مدل پیش‌بینی است چرا که حذف پارامترهای نامربوط یا زائد می‌تواند تفسیرپذیری و عمومیت‌پذیری مدل را بهبود بخشد. در این پژوهش از دو روش فیلتر و لاسو[21] برای تعیین پارامترهای مؤثر استفاده شد. شکل (۴) و شکل (۵) نشان‌دهنده ماتریس همبستگی پارامترهای انتخاب شده به ترتیب در خوشه‌های ۱ و ۲ با روش فیلتر می‌باشند. در روش فیلتر همبستگی خطی بین هر پارامتر و پارامتر هدف اندازه‌گیری شده و پارامترها بر اساس مقادیر همبستگی آن‌ها با هدف، امتیازدهی می‌شوند. بدین ترتیب با در نظر گرفتن حد آستانه ۰/۴ برای ضریب همبستگی، پارامترهایی که همبستگی کمتر از این مقدار با پارامتر هدف دارند، حذف می‌شوند. در ادامه برای جلوگیری از افزونگی، هر پارامتری که همبستگی بالای ۰/۷ با پارامترهای دیگر دارد نیز از لیست خارج می‌شود. اما در روش لاسو به هر یک از پارامترها ضرایبی اختصاص داده می‌شود و بر اساس آن‌ها سعی به حذف پارامترهای کم‌اهمیت‌تر و انتخاب پارامترهای مهم‌تر می‌شود. این موضوع بدین صورت انجام می‌پذیرد که به کم‌اهمیت‌ترین پارامترها ضریب صفر داده شده و از مدل حذف می‌گردند و تنها پارامترهای با ضرایب غیر صفر باقی می‌مانند. لازم به ذکر است که درجه سخت‌گیری مدل در روش لاسو به کمک پارامتر تنظیم λ صورت می‌پذیرد. مقدار بالاتر λ منجر به منظم‌سازی قوی‌تر و در نتیجه نسبت دادن تعداد ضرایب بیشتری به عدد صفر می‌شود. این در حالی است

که، مقدار کمتر λ میزان منظم شدن را کاهش داده و به ضرایب بیشتری اجازه می‌دهد که مقادیر غیر صفر داشته باشند. در این پژوهش به کمک روش جست‌وجوی شبکه‌ای بازه ۰/۰۱ تا ۱ به ۱۰۰ قسمت تقسیم شده و بهینه‌ترین حالت برای λ به دست آمده است. لازم به ذکر است که مقادیر کمتر از ۰/۰۱ برای λ منجر به انتخاب تعداد بسیار زیادی از پارامترها خواهد شد که به هیچ عنوان بهینه نیست. مقدار λ در این پژوهش برابر با ۰/۰۱ به دست آمده است. نتایج نهایی پارامترهای انتخاب شده در جدول (۴) قابل مشاهده است. لازم به ذکر است که در این جدول، Pre بیانگر پارامتر بارندگی و LST بیانگر پارامتر دمای سطح زمین می‌باشند. ضمنا پسوندهای ۱ و ۳ به ترتیب بیانگر میانگین یک ماهه و میانگین متحرک سه ماهه آن پارامتر می‌باشند.



شکل ۴ ماتریس همبستگی پارامترهای انتخاب شده در خوشه ۱ با روش فیلتر



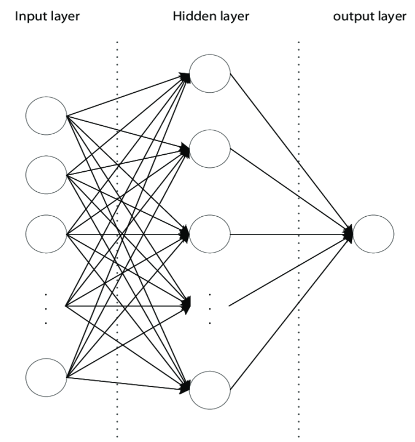
شکل ۵ ماتریس همبستگی پارامترهای انتخاب شده در خوشه ۲ با روش فیلتر

جدول ۶ نتایج انتخاب پارامترهای مؤثر

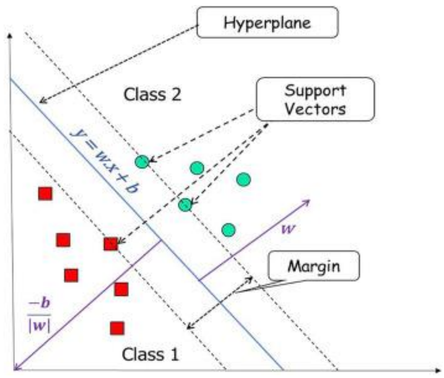
نام روش	خوشه	تعداد پارامترهای انتخاب شده	پارامترهای انتخاب شده
فیلتر	شماره ۱	۷	LST_night-3, NDDI6-1, NDVI-3, NDDI5-1, NDWI5-3, Pre-1, Pre-3
	شماره ۲	۶	LST_mean-3, NDDI7-1, NDDI6-3, Pre-1, NMDI-1, Pre-3
لاسو	شماره ۱	۶	Pre-3, LST_day-1, LST_mean-3, LST_night-3, NDDI6-1, NDDI7-3
	شماره ۲	۷	Pre-3, LST_day-1, LST_day-3, LST_mean-3, LST_night-3, NDDI7-3, NDVI-3

پرسپترون چندلایه

پرسپترون چند لایه (MLP) نوعی شبکه عصبی مصنوعی است که اولین بار توسط رازنبالت [22] معرفی گردید. MLPها که از چندین لایه گره به هم پیوسته به نام نورون تشکیل شده‌اند، به نوعی شبکه عصبی پیشخور به حساب می‌آیند، به این معنی که اطلاعات در یک جهت و ابتدا از لایه ورودی و سپس از طریق یک یا چند لایه پنهان به لایه خروجی جریان می‌یابند. به طور کلی شبکه‌های عصبی مصنوعی با پردازش داده‌های تجربی قانون نهفته بین متغیرهای یک مدل را شناسایی کرده و به نحوی مناسب و با سرعت بالا به ساختار شبکه انتقال می‌دهند[23]. در این شبکه‌ها هر نورون سیگنال‌های ورودی را دریافت کرده، یک تابع فعال‌سازی غیرخطی را به مجموع وزنی ورودی‌ها اعمال کرده و خروجی را به لایه بعدی ارسال می‌کند. ساختار کلی این شبکه‌ها در شکل (۶) مشاهده می‌شود. به طور کلی MLPها قادر به یادگیری روابط پیچیده غیرخطی هستند و به طور گسترده برای کارهایی مانند طبقه‌بندی، رگرسیون و تشخیص الگو مورد استفاده قرار می‌گیرند. در این پژوهش تعداد لایه‌های پنهان برابر با ۱ و تعداد نورون‌های لایه پنهان برابر با ۱ می‌باشند. ضمنا از واحد خطی اصلاح شده به عنوان تابع فعال‌سازی استفاده شده است.



شکل ۶ ساختار شبکه پرسپترون چند لایه[32]



شکل ۷ ساختار الگوریتم ماشین بردار پشتیبان[33]

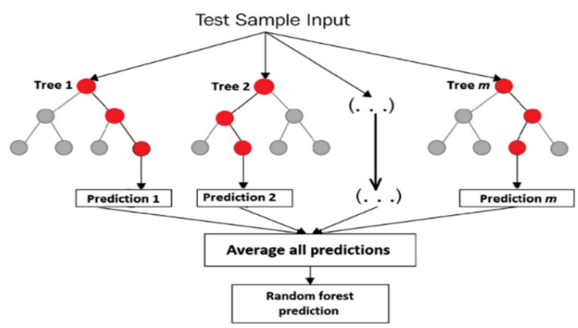
ماشین بردار پشتیبان

ماشین بردار پشتیبان (SVM) که اولین بار توسط کورتس و واپنیک [24] مورد استفاده قرار گرفت، یک الگوریتم یادگیری ماشین قدرتمند و پرکاربرد است و زیرمجموعه یادگیری نظارت شده به حساب می‌آید. این الگوریتم ابتدا داده‌های ورودی و خروجی را دریافت کرده و سپس اگر داده‌ها به صورت خطی در فضای ویژگی اصلی مسئله قابل تفکیک نباشند، از یک تابع کرنل برای تبدیل داده‌ها به یک فضای ویژگی با ابعاد بالاتر استفاده می‌کند. هدف از این تبدیل یافتن یک رابطه خطی بین متغیرهای ورودی و مقادیر هدف است. به همین منظور در مدل‌سازی رگرسیونی، SVM به یافتن ابرصفحه‌ای می‌پردازد که بهترین تطابق با داده‌ها را داشته باشد و در عین حال خطاهای پیش‌بینی را به حداقل برساند. این ابرصفحه با تنظیم پارامترهای مدل برای به حداقل رساندن مجموع مقادیر خطا بین مقادیر هدف پیش‌بینی شده و واقعی تعیین می‌شود. اصطلاح بردارهای پشتیبان، اشاره به نزدیک‌ترین نقاط داده به ابرصفحه را دارد که این نقاط بر موقعیت و جهت ابرصفحه تأثیر گذاشته و در تعیین پیش‌بینی‌های مدل نقش دارند. هنگامی که ابرصفحه بهینه تعیین شد، می‌توان از آن برای پیش‌بینی مقادیر هدف و نیز برای داده‌های جدید و نادیده استفاده کرد. مقدار هدف پیش‌بینی شده برای داده‌های جدید بر اساس موقعیت آن‌ها نسبت به ابرصفحه

محاسبه خواهد شد. ساختار این الگوریتم در شکل (۷) نمایش داده شده است. در این پژوهش از تابع پایه شعاعی به عنوان تابع هسته استفاده شده و مقیاس آن ۲/۶ در نظر گرفته شده است.

جنگل تصادفی

الگوریتم جنگل تصادفی (RF) که اولین بار توسط بریمن[25] مورد استفاده قرار گرفت، یک الگوریتم یادگیری ماشین قدرتمند است که می‌تواند برای مسئله‌های رگرسیونی مورد استفاده قرار گیرد. این الگوریتم یک روش یادگیری گروهی است که برای پیش‌بینی، چندین درخت تصمیم را ترکیب می‌کند. جنگل تصادفی پس از دریافت داده‌های ورودی و متغیر هدف، مجموعه‌ای از درختان تصمیم را می‌سازد که در آن هر درخت تصمیم با استفاده از یک زیرمجموعه تصادفی از مجموعه داده‌های مربوط به آموزش ساخته می‌شود. علاوه بر این، یک زیرمجموعه تصادفی از پارامترها در هر گره از درخت انتخاب می‌شود تا بهترین تقسیم را انجام داده و تنوع را در بین درختان فراهم کند. سپس هر درخت تصمیم به صورت بازگشتی و با پارتیشن‌بندی داده‌ها در هر گره بر اساس یک معیار تقسیم انتخابی، که معمولا کاهش واریانس یا میانگین مربعات خطاست، رشد می‌کند. این تقسیم تا زمانی که حداقل یکی از معیارهای توقف برآورده شود، ادامه می‌یابد. این معیارها می‌توانند شامل رسیدن به حداکثر عمق یا حداقل تعداد نمونه‌های مورد نیاز برای یک گره برگ باشند. در ادامه، الگوریتم با ترکیب پیش‌بینی‌های همه درخت‌های تصمیم، یک مجموعه ایجاد می‌کند. در رگرسیون، پیش‌بینی‌های فردی از هر درخت برای به دست آوردن پیش‌بینی نهایی میانگین‌گیری می‌شوند. این فرایند میانگین‌گیری به کاهش تأثیر نقاط پرت و نویز در داده‌ها کمک می‌کند. نحوه عملکرد این الگوریتم در شکل (۸) قابل مشاهده است.



شکل ۸ نحوه عملکرد الگوریتم جنگل تصادفی [34]

جنگل تصادفی تصحیح بایاس شده

همان‌طور که پیش‌تر نیز اشاره شد، در الگوریتم جنگل تصادفی هر درخت تصمیم در مجموعه درختان، پیش‌بینی‌های مستقلی انجام می‌دهد و پیش‌بینی نهایی با میانگین‌گیری پیش‌بینی‌های همه درختان به دست می‌آید. با این حال، با وجود اینکه این میانگین‌گیری به نحوی از تولید مقادیر پیش‌بینی شده غیرمنطقی جلوگیری می‌کند، ممکن است پیش‌بینی‌های این الگوریتم، به‌ویژه در شرایطی که به عنوان مثال داده‌های آموزشی ناهمگون هستند یا روابط غیرخطی در داده‌ها وجود دارد و یا زمانی که مجموعه مشاهدات بسیار بزرگ یا بسیار کوچک هستند، مقادیری بایاس از خود نشان دهند [26].

الگوریتم جنگل تصادفی تصحیح بایاس شده (BRF) به منظور در نظر گرفتن همین اثر، توسط ژنگ و لو [27] معرفی شده است. این الگوریتم ابتدا مطابق با رابطه (۲) یک مدل RF ساده را بر روی داده‌های آموزشی اعمال کرده و نتایج پیش‌بینی شده را به‌دست می‌آورد.

$$Y_{train} = RF(X_{train})$$

Error! No text of specified style in document.)

در این رابطه، X_{train} بیانگر متغیرهای مستقل داده‌های آموزش و Y_{train} بیانگر متغیر وابسته داده‌های آموزش می‌باشند. در ادامه مقادیر باقی‌مانده با توجه به مقادیر پیش‌بینی شده مدل RF توسط رابطه (۳) محاسبه می‌گردد.

$$r_{train} = Y_{train} - \hat{Y}_{train}$$

Error! No text of specified style in document.)

در این رابطه، r_{train} بیانگر باقی‌مانده داده‌های آموزش و $\hat{Y}_{train}$ بیانگر مقادیر پیش‌بینی شده متغیر وابسته توسط RF می‌باشند.

در ادامه مطابق با رابطه (۴)، یک مدل RF دیگر به باقی‌مانده‌ها به عنوان متغیر وابسته و کل داده‌های آموزشی به عنوان متغیرهای مستقل برازش داده می‌شود.

$$r_{train} = RF_{res}(X_{train}, Y_{train}) \quad (۴)$$

سپس مقادیر پیش‌بینی شده متغیر وابسته داده‌های آزمون توسط همان مدل RF استفاده شده در رابطه (۲)، مطابق رابطه (۵) تولید می‌شوند.

$$Y_{test} = RF(X_{test}) \quad (۵)$$

در این رابطه، X_{test} بیانگر متغیرهای مستقل داده‌های آزمون و Y_{test} بیانگر متغیر وابسته داده‌های آزمون می‌باشند.

سپس درست به مانند گام سوم، مقادیر باقی‌مانده داده‌های آزمون توسط همان مدل RF موجود در گام سوم، و مطابق رابطه (۶) محاسبه می‌شوند.

$$r_{test} = RF_{res}(X_{test}, Y_{test}) \quad (۶)$$

در این رابطه، r_{test} بیانگر باقی‌مانده داده‌های آزمون می‌باشد.

در انتها مقادیر متغیر وابسته به صورت اصلاح بایاس شده، از طریق رابطه (۷) قابل محاسبه خواهند بود.

$$Y_{bias-corrected} = Y_{test} + r_{test} \quad (۷)$$

در این رابطه، $Y_{bias-corrected}$ بیانگر متغیر وابسته به صورت اصلاح بایاس شده ، Y_{test} بیانگر متغیر وابسته داده‌های آزمون و r_{test} بیانگر باقی‌مانده داده‌های آزمون می‌باشند. در این پژوهش، تعداد درختان تصمیم در هر دو الگوریتم RF و BRF ۵۰۰ در نظر گرفته شده است.

اعتبارسنجی مدل

هدف نهایی در مدل‌سازی با الگوریتم‌های یادگیری ماشین، توسعه مدلی است که بتواند پیش‌بینی‌های دقیقی بر روی داده‌های جدید و دیده نشده انجام دهد. لذا به هنگام فرایند آموزش مدل، ارزیابی عملکرد آن مدل و نیز اطمینان حاصل کردن از توانایی تعمیم مدل به داده‌هایی فراتر از داده‌های آموزشی، امری ضروری به حساب می‌آید. برای این منظور از مفهوم اعتبارسنجی استفاده می‌شود. یکی از پرکاربردترین روش‌های اعتبارسنجی، روش اعتبارسنجی متقابل K-fold می‌باشد. در پژوهش حاضر، از این روش برای اعتبارسنجی تمامی الگوریتم‌های استفاده شده، بهره گرفته شده است. ضمنا در تمامی الگوریتم‌ها K=5 در نظر گرفته شده است.

معیارهای ارزیابی عملکرد مدل

در هر فرایند مدل‌سازی، عملکرد نهایی یک مدل برای پیش‌بینی متغیر هدف نیاز به ارزیابی دارد. این وظیفه بر عهده معیارهای ارزیابی است. این معیارها، روش‌های عددی هستند که بسته به نوع مسئله، میزان عملکرد مدل را از نظر دقت، یادآوری، خطا یا سایر معیارهای مرتبط ارزیابی کرده و امکان انتخاب بهترین و دقیق‌ترین مدل را فراهم می‌آورند. در این پژوهش از معیارهای ارزیابی MSE، RMSE، MAE و R^2 برای ارزیابی عملکرد مدل استفاده شده است. روابط محاسبه این معیارها در ادامه قابل مشاهده است.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \quad (۸)$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \quad (۹)$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \quad (۱۰)$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \quad (۱۱)$$

در این روابط n تعداد داده‌ها، y_i مقادیر واقعی، $\hat{y}_i$ مقادیر پیش‌بینی شده و $\bar{y}$ میانگین مقادیر واقعی می‌باشند.

نتایج و بحث

ارزیابی عملکرد مدل‌ها

پس از اعمال پردازش‌های مربوطه بر روی داده‌ها و نیز تنظیم فراپارامترهای الگوریتم‌های یادگیری ماشین، فرایند مدل‌سازی با استفاده از چهار الگوریتم یاد شده و با توجه به نتایج انتخاب پارامترهای مؤثر صورت گرفت. نتایج این مدل‌سازی برای خوشه ۱ در جدول (۵) و برای خوشه ۲ در جدول (۶) قابل مشاهده است. لازم به ذکر است که فرایند مدل‌سازی برای هر یک از ترکیب‌های پارامترهای ورودی به تعداد ۱۰ مرتبه تکرار شد و میانگین نتایج برای هر مدل در جدول‌های مربوطه ثبت گردید.

همان طور که از جدول (۵) و جدول (۶) پیداست، عملکرد تمامی مدل‌ها به طور قابل توجهی بالا می‌باشد. این بدین معنی است که متغیرهای مستقل (پارامترهای ورودی) به

شدت با متغیر وابسته (پارامتر هدف) مرتبط بوده و می‌توانند به طور مؤثر مقادیر آن را پیش‌بینی کنند. به عنوان مثال، همان طور که از ماتریس‌های همبستگی روش فیلتر در شکل‌های (۴) و (۵) پیداست، ارتباط و وابستگی میان برخی از پارامترهای ورودی با پارامتر هدف به طور قابل توجهی بالاست. مثلا برای خوشه شماره (۱) پارامتر میانگین متحرک سه ماهه دمای سطح زمین در شب (LST_night-3) و برای خوشه شماره (۲) پارامتر میانگین متحرک سه ماهه میانگین دمای سطح زمین در روز و شب (LST_mean-3) همبستگی بالای ۹۰ درصد با شاخص SPEI به عنوان پارامتر هدف از خود نشان داده‌اند.

با این حال، می‌توان از نتایج به دست آمده این گونه استنباط کرد که برای ایستگاه‌های متعلق به خوشه ۱، الگوریتم جنگل تصادفی تصحیح بایاس شده با ثبت بالاترین مقدار در معیار ضریب تعیین ($R^2 = 0.9858$) و پایین‌ترین مقدار در معیار ریشه میانگین مربعات خطا ($RMSE = 0.1190$) بهترین عملکرد را برای پیش‌بینی شاخص SPEI و در نتیجه پیش‌بینی خشکسالی کشاورزی داراست. پس از الگوریتم جنگل تصادفی تصحیح بایاس شده، به ترتیب الگوریتم‌های جنگل تصادفی، پرسپترون چندلایه و ماشین بردار پشتیبان بهترین عملکرد را از خود نشان دادند. جزئیات عملکرد و نتایج این الگوریتم‌ها در جدول (۵) قابل مشاهده است.

همچنین به مانند خوشه ۱، برای ایستگاه‌های متعلق به خوشه ۲ نیز الگوریتم جنگل تصادفی تصحیح بایاس شده با ثبت بالاترین مقدار در معیار ضریب تعیین ($R^2 = 0.9809$) و پایین‌ترین مقدار در معیار ریشه میانگین مربعات خطا ($RMSE = 0.1375$) بهترین عملکرد را برای پیش‌بینی شاخص SPEI و در نتیجه پیش‌بینی خشکسالی کشاورزی از خود نشان داد. در این خوشه پس از الگوریتم جنگل تصادفی تصحیح بایاس شده، به ترتیب الگوریتم‌های ماشین بردار پشتیبان، جنگل تصادفی و پرسپترون چندلایه بهترین عملکرد را از خود به جای گذاشتند. جزئیات عملکرد و نتایج این الگوریتم‌ها در جدول (۶) قابل مشاهده است.

جدول ۷ نتایج مدل‌سازی پیش‌بینی خشکسالی در خوشه ۱ به کمک الگوریتم‌های یادگیری ماشین

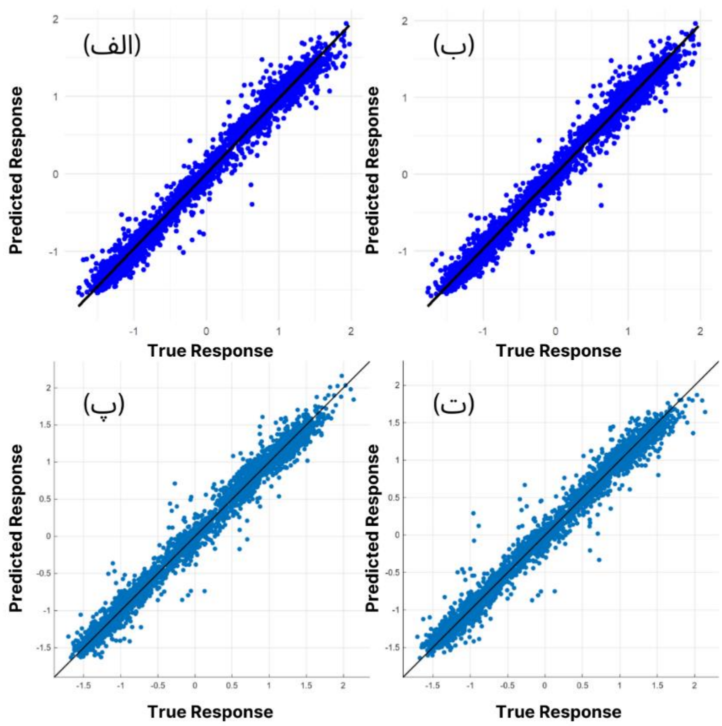
نام الگوریتم یادگیری ماشین	روش انتخاب پارامتر مؤثر	R^2	RMSE	MSE	MAE	زمان آموزش مدل (ثانیه)
پرسپترون چندلایه	فیلتر	۰/۹۸۳۴	۰/۱۲۸۶	۰/۰۱۶۵	۰/۰۸۹۴	۴۰
	لاسو	۰/۹۸۵۲	۰/۱۲۱۰	۰/۰۱۴۷	۰/۰۸۵۳	۳۸
ماشین بردار پشتیبان	فیلتر	۰/۹۸۲۲	۰/۱۳۳۰	۰/۰۱۷۷	۰/۰۹۱۵	۱۹
	لاسو	۰/۹۸۴۸	۰/۱۲۲۹	۰/۰۱۵۱	۰/۰۸۳۹	۱۲
جنگل تصادفی	فیلتر	۰/۹۸۳۲	۰/۱۳۰۲	۰/۰۱۷۰	۰/۰۹۱۱	۲۰۴
	لاسو	۰/۹۸۵۸	۰/۱۱۹۱	۰/۰۱۴۲	۰/۰۸۲۰	۲۰۵
جنگل تصادفی تصحیح بایاس شده	فیلتر	۰/۹۸۴۱	۰/۱۲۶۳	۰/۰۱۶۰	۰/۰۸۷۹	۲۱۴
	لاسو	۰/۹۸۵۸	۰/۱۱۹۰	۰/۰۱۴۲	۰/۰۸۲۱	۲۱۸

جدول ۸ نتایج مدل‌سازی پیش‌بینی خشکسالی در خوشه ۲ به کمک الگوریتم‌های یادگیری ماشین

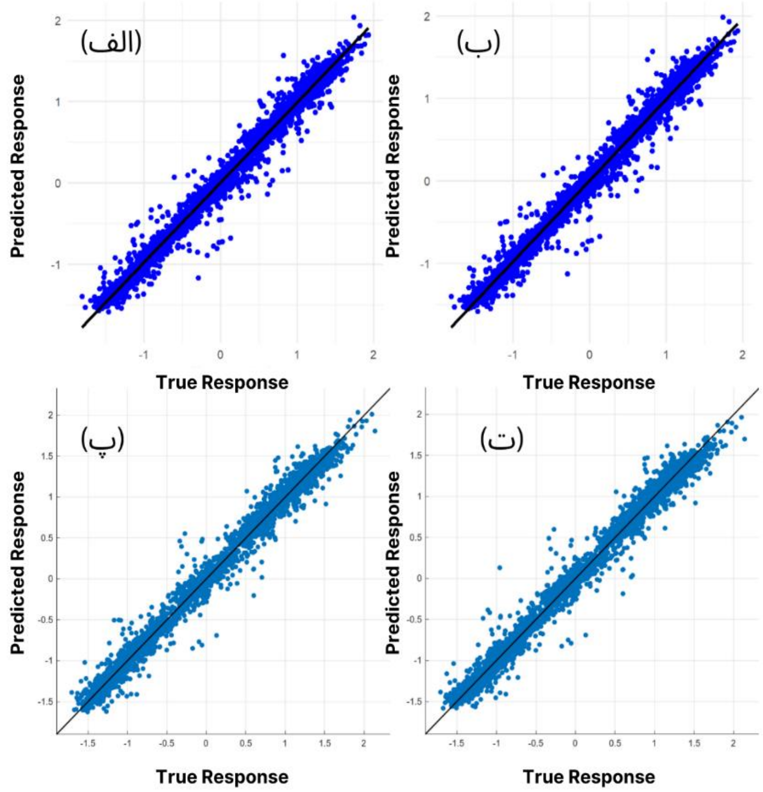
نام الگوریتم یادگیری ماشین	روش انتخاب پارامتر مؤثر	R^2	RMSE	MSE	MAE	زمان آموزش مدل (ثانیه)
پرسپترون چندلایه	فیلتر	۰/۹۷۷۰	۰/۱۵۱۲	۰/۰۲۲۹	۰/۱۱۴۱	۵
	لاسو	۰/۹۷۸۶	۰/۱۴۵۷	۰/۰۲۱۳	۰/۱۰۸۸	۶
ماشین بردار پشتیبان	فیلتر	۰/۹۷۶۳	۰/۱۵۳۴	۰/۰۲۳۵	۰/۱۱۴۴	۲
	لاسو	۰/۹۸۰۳	۰/۱۴۰۱	۰/۰۱۹۷	۰/۱۰۴۵	۲
جنگل تصادفی	فیلتر	۰/۹۷۶۱	۰/۱۵۷۸	۰/۰۲۵۰	۰/۱۱۶۴	۳
	لاسو	۰/۹۸۰۰	۰/۱۴۰۴	۰/۰۱۹۸	۰/۱۰۵۹	۳
جنگل تصادفی تصحیح بایاس شده	فیلتر	۰/۹۷۷۹	۰/۱۴۹۶	۰/۰۲۲۵	۰/۱۱۰۹	۳
	لاسو	۰/۹۸۰۹	۰/۱۳۷۵	۰/۰۱۸۹	۰/۱۰۳۱	۴

همان طور که از نتایج نیز پیداست، به طور کلی می‌توان عملکرد الگوریتم جنگل تصادفی را در مقایسه با ماشین بردار پشتیبان و پرسپترون چندلایه بهتر دانست. این امر به این دلیل می‌باشد که جنگل تصادفی دارای چندین مزیت بوده که آن را در چنین سبک مسائل به انتخابی بهتر تبدیل می‌کند. به عنوان مثال این الگوریتم به دلیل استفاده از چندین درخت تصمیم‌گیری و بوت‌استرپ معمولا نسبت به دو الگوریتم دیگر مقاومت بیشتری در برابر بیش‌برازش دارد. همچنین می‌توان به تفسیرپذیری بهتر این الگوریتم نسبت به دو الگوریتم دیگر اشاره کرد. با این حال، جنگل تصادفی معمولا در بازتولید مشاهدات شدید عملکرد ضعیفی دارد. به طوری که گاه مقادیر بزرگ کمتر از حد و مقادیر کوچک بیشتر از حد برآورد می‌شوند [25]. لذا از آن‌جایی که خشکسالی معمولا به عنوان شرایط خشک بیش از اندازه و غیرعادی تلقی می‌شود، استفاده از جنگل تصادفی ممکن است منجر به ایجاد بایاس در فرایند پایش خشکسالی گردد. به همین جهت، ترکیب روش تصحیح بایاس با الگوریتم جنگل تصادفی می‌تواند به حذف بایاس ایجاد شده و حصول نتایج بهتر در مدل‌سازی منجر گردد. دستیابی به مقادیر ضریب تعیین بالاتر و ریشه میانگین مربعات خطای پایین‌تر برای روش جنگل تصادفی تصحیح بایاس شده نسبت به جنگل تصادفی ساده، مهر تأییدی بر این ادعا در این پژوهش می‌باشد.

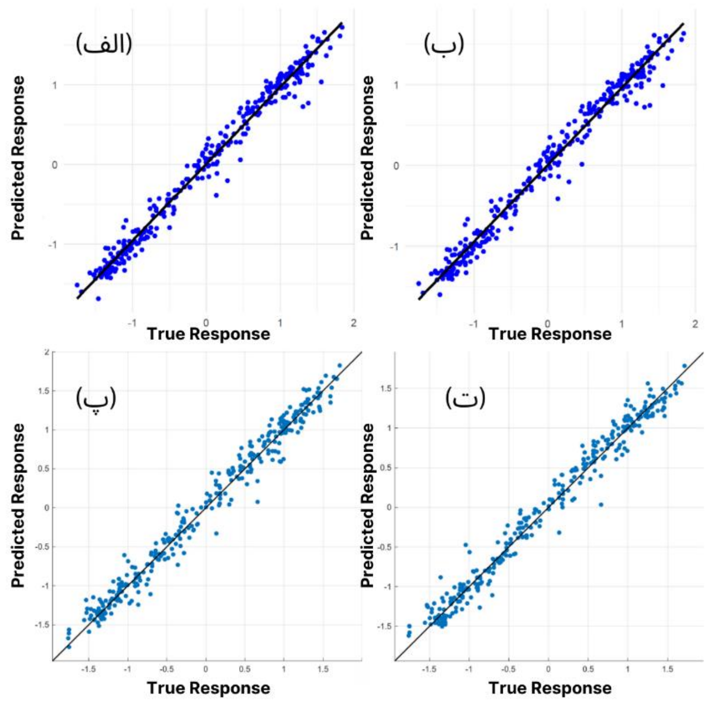
در ادامه شکل (۹) و شکل (۱۰) نشان‌دهنده مقایسه بین مقادیر پیش‌بینی شده و مقادیر واقعی SPEI در خوشه ۱ و شکل (۱۱) و شکل (۱۲) بیانگر مقایسه بین مقادیر پیش‌بینی شده و مقادیر واقعی SPEI در خوشه ۲ می‌باشند. همچنین شکل (۱۳) نشان‌دهنده گراف مقایسه نتایج مدل‌ها بر اساس معیار ضریب تعیین، شکل (۱۴) نشان‌دهنده گراف مقایسه نتایج مدل‌ها بر اساس معیار ریشه میانگین مربعات خطا، شکل (۱۵) نشان‌دهنده گراف مقایسه نتایج مدل‌ها بر اساس معیار میانگین مربعات خطا و شکل (۱۶) نشان‌دهنده گراف مقایسه نتایج مدل‌ها بر اساس معیار میانگین خطای مطلق می‌باشند.



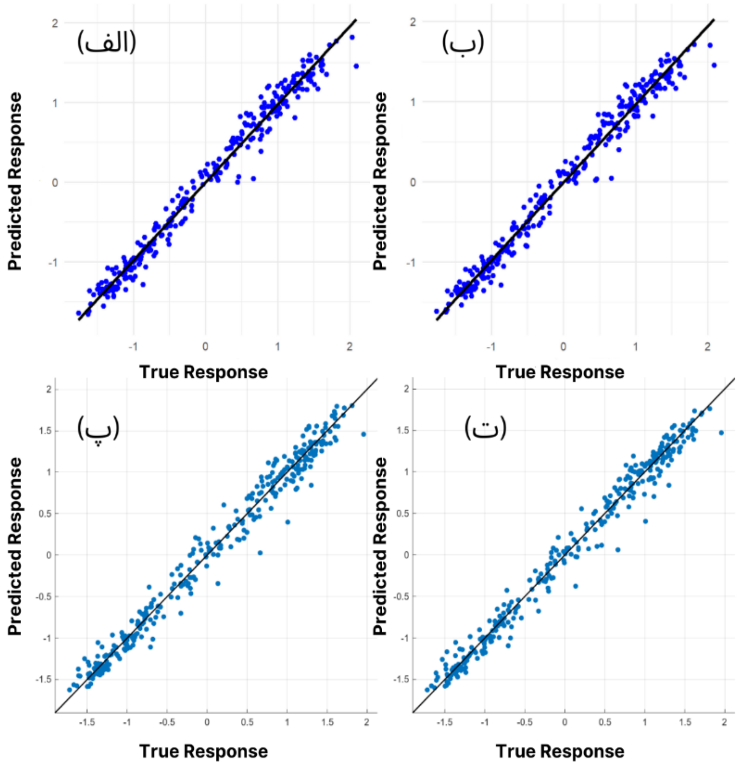
شکل ۹ مقایسه مقادیر پیش‌بینی شده و واقعی SPEI با روش فیلتر در خوشه ۱: الف) الگوریتم جنگل تصادفی تصحیح بایاس شده، ب) الگوریتم جنگل تصادفی، پ) الگوریتم پرسپترون چندلایه و ت) الگوریتم ماشین بردار پشتیبان



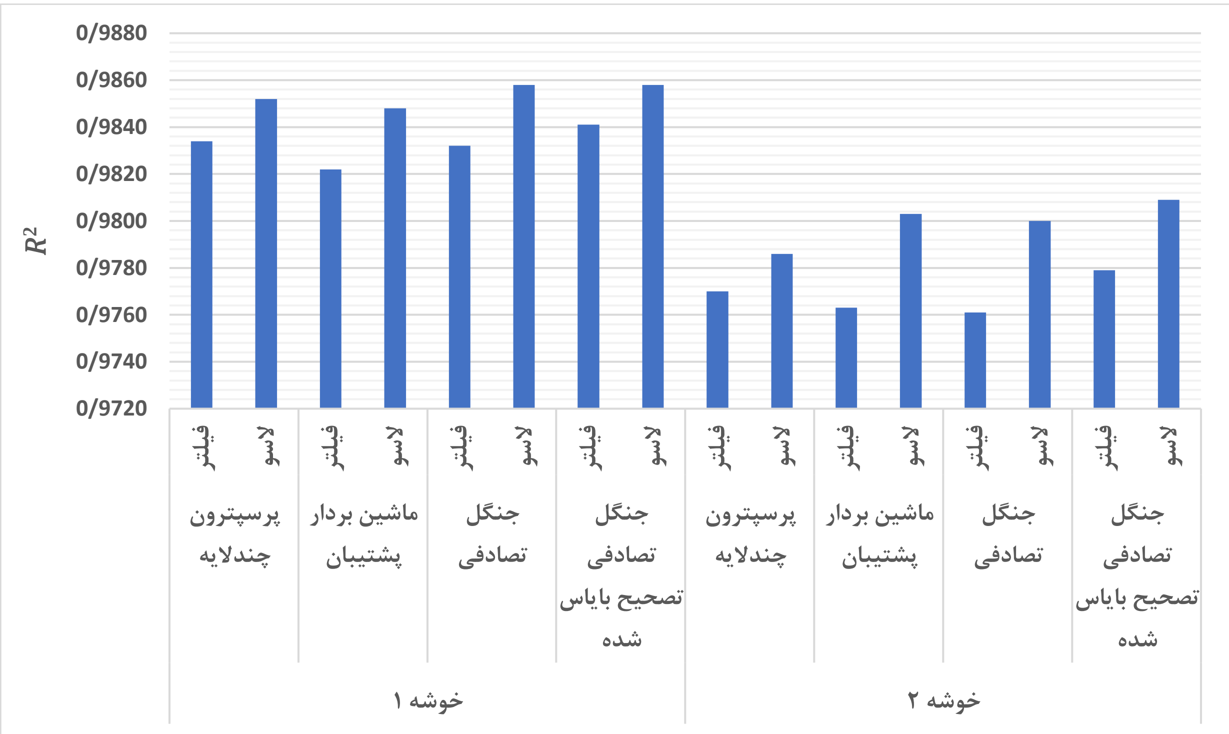
شکل ۱۰ مقایسه مقادیر پیش‌بینی شده و واقعی SPEI با روش لاسو در خوشه ۱: الف) الگوریتم جنگل تصادفی تصحیح بایاس شده، ب) الگوریتم جنگل تصادفی، پ) الگوریتم پرسپترون چندلایه و ت) الگوریتم ماشین بردار پشتیبان



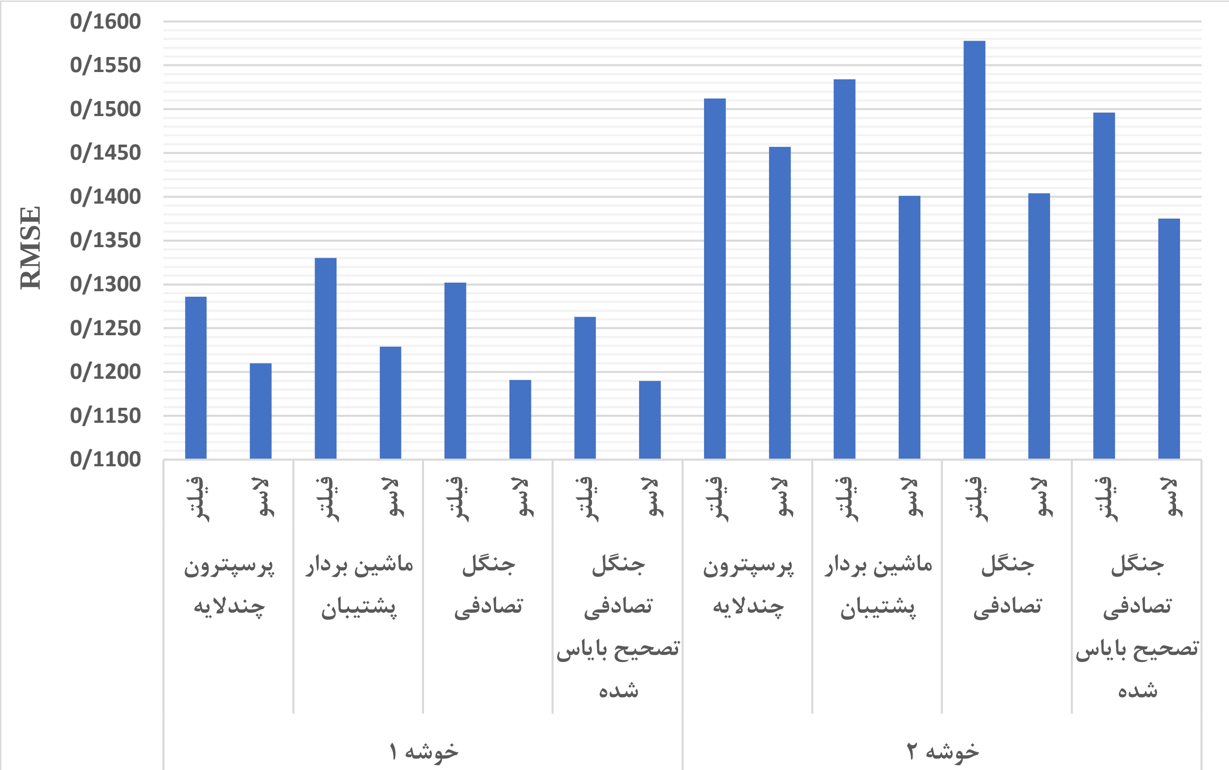
شکل ۱۱ مقایسه مقادیر پیش‌بینی شده و واقعی SPEI با روش فیلتر در خوشه ۲: الف) الگوریتم جنگل تصادفی تصحیح بایاس شده، ب) الگوریتم جنگل تصادفی، پ) الگوریتم پرسپترون چندلایه و ت) الگوریتم ماشین بردار پشتیبان



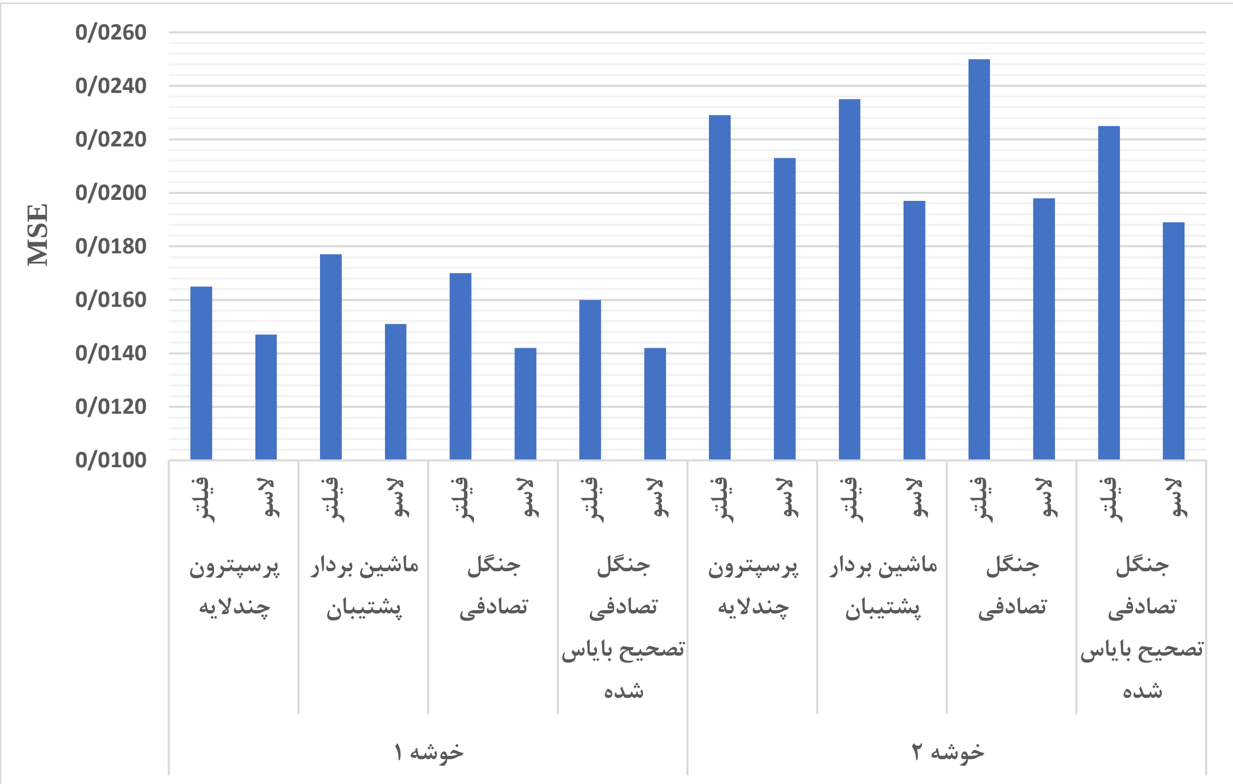
شکل ۱۲ مقایسه مقادیر پیش‌بینی شده و واقعی SPEI با روش لاسو در خوشه ۲: الف) الگوریتم جنگل تصادفی تصحیح بایاس شده، ب) الگوریتم جنگل تصادفی، پ) الگوریتم پرسپترون چندلایه و ت) الگوریتم ماشین بردار پشتیبان



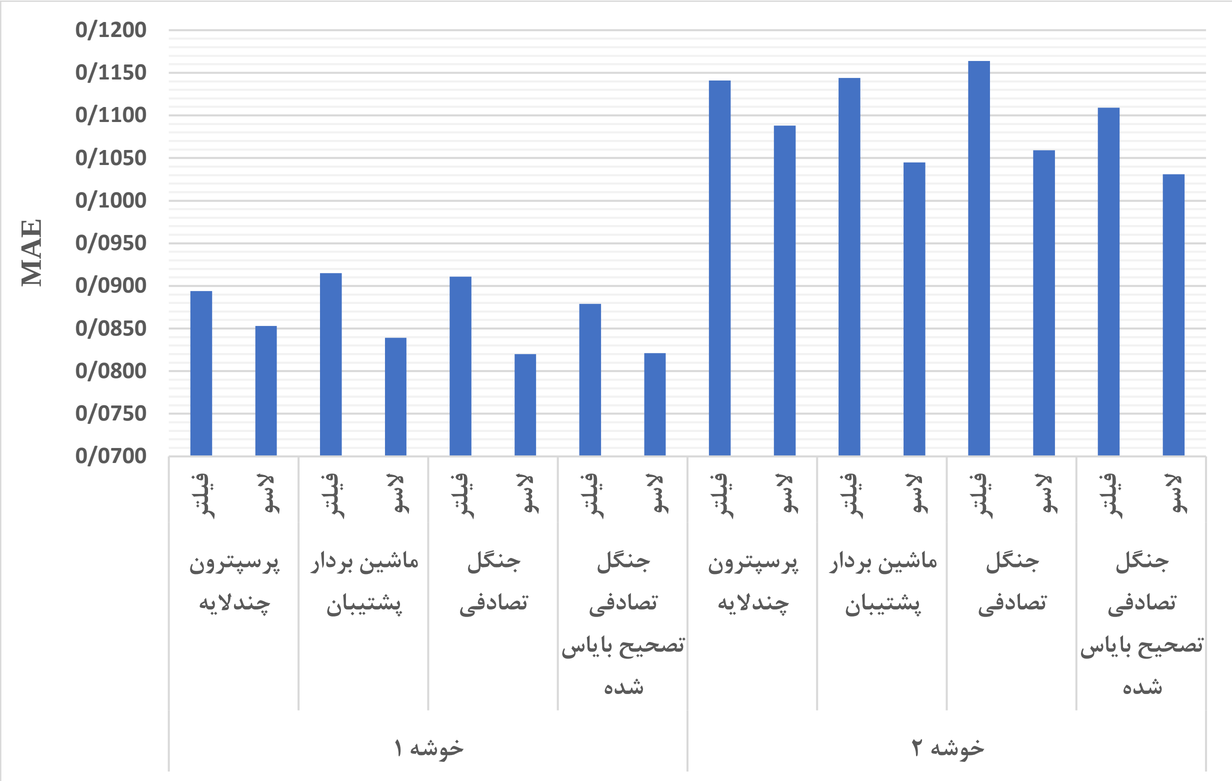
شکل ۱۳ مقایسه نتایج مدل‌ها بر اساس معیار ضریب تعیین



شکل ۱۴ مقایسه نتایج مدل‌ها بر اساس معیار ریشه میانگین مربعات خطا



شکل ۱۵ مقایسه نتایج مدل‌ها بر اساس معیار میانگین مربعات خطا



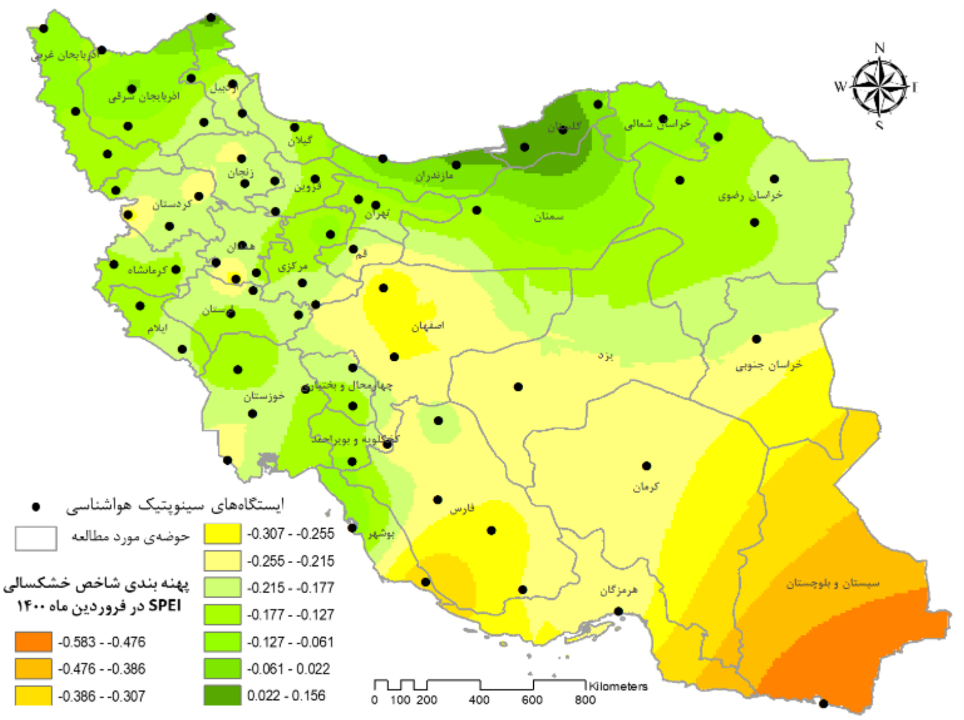
شکل ۱۶ مقایسه نتایج مدل‌ها بر اساس معیار میانگین خطای مطلق

تهیه نقشه‌های پهنه‌بندی خشکسالی کشاورزی

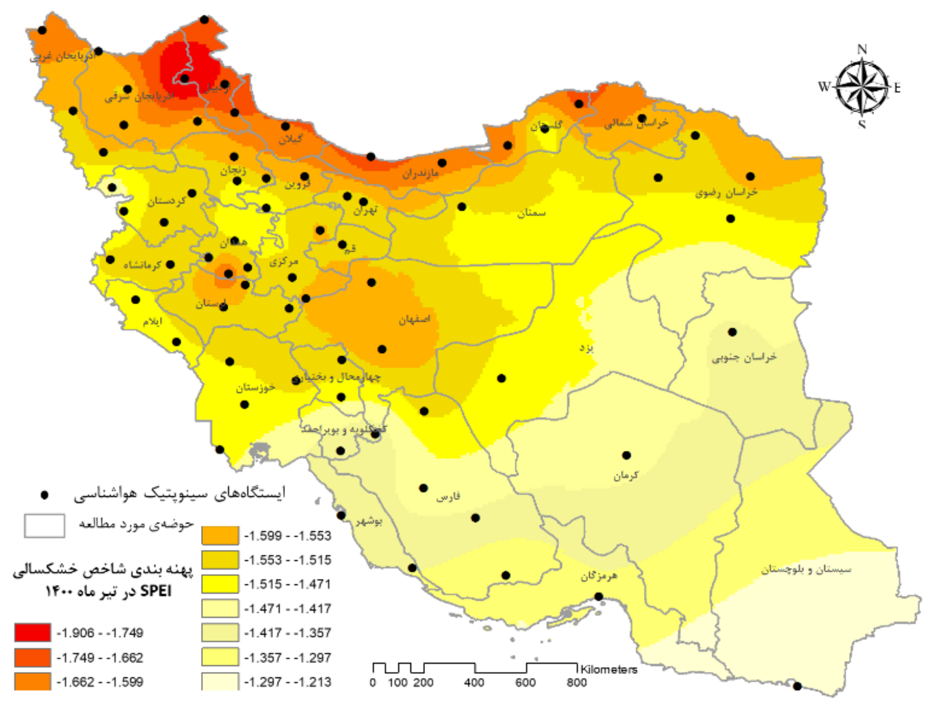
پس از ایجاد مدل‌های پیش‌بینی خشکسالی و انتخاب مدل با بهترین عملکرد، می‌توان از نتایج این مدل برای تهیه نقشه‌های پهنه‌بندی خشکسالی استفاده کرد. این نقشه‌ها ابزاری هستند که برای شناسایی و طبقه‌بندی مناطق مستعد به شرایط خشکسالی مورد استفاده قرار می‌گیرند. برای تهیه این نقشه‌ها، از مقادیر پیش‌بینی شده مدل ایجاد شده با الگوریتم جنگل تصادفی تصحیح بایاس شده برای شاخص SPEI، استفاده شده و این مقادیر در چهار ماه منتخب از سال که هر کدام نمونه‌ای برای یکی از چهار فصل سال می‌باشند، در تمامی ایستگاه‌های مورد مطالعه محاسبه شده و پس از آن بر روی سطح کل کشور درون‌یابی می‌شوند.

همان طور که پیش‌تر نیز مطرح شد، عموما خشکسالی به حالتی اطلاق می‌شود که شاخص SPEI مقادیری کمتر از ۱- داشته باشد. این در حالی است که مقادیر بزرگ‌تر از یک برای این شاخص نشان‌دهنده شرایط مرطوب، مقادیر بین صفر و منفی یک نشان‌دهنده شرایط نسبتا خشک و مقادیر بین صفر و یک نشان‌دهنده شرایط نسبتا مرطوب می‌باشند. به این ترتیب و با توجه به شکل (۱۷) که نشان‌دهنده نقشه پهنه‌بندی خشکسالی کشاورزی در ایران برای فروردین ۱۴۰۰ می‌باشد، مشاهده می‌شود که مدل پیش‌بینی خشکسالی برای اغلب مناطق کشور حالت نسبتا خشک را پیش‌بینی کرده و تنها بخش‌هایی از استان‌های گلستان و مازندران را به حالت نسبتا مرطوب نسبت داده است. در ادامه همان طور که از شکل (۱۸) پیداست، مدل برای تمام مناطق کشور در تیر ماه ۱۴۰۰ شرایط خشکسالی را پیش‌بینی کرده است. بررسی دقیق‌تر این نقشه نشان می‌دهد که قسمت‌هایی از استان‌های مازندران، گیلان، اردبیل و آذربایجان شرقی، که همگی از استان‌های پربارش کشور به حساب می‌آیند، تحت شرایط خشکسالی شدیدتری نسبت به سایر مناطق کشور قرار دارند. این بدین خاطر است که به علت برقراری شرایط مرطوب در سایر اوقات سال در این مناطق، کاهش نسبی میزان بارندگی و گرمای هوا در تابستان منجر به افت شدیدتر مقادیر شاخص SPEI نسبت به سایر مناطق در این فصل می‌گردد. شکل (۱۹) نیز به مانند شکل (۱۷) نشان‌دهنده حالت نسبتا خشک برای اغلب مناطق کشور و حالت نسبتا مرطوب برای تنها بخش‌هایی از استان‌های مازندران و گیلان در

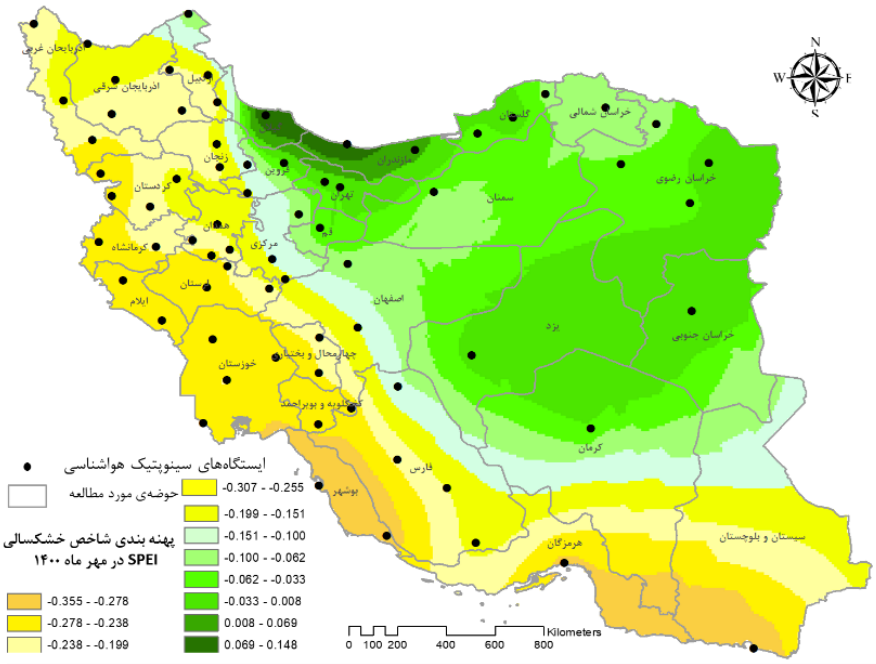
مهر ۱۴۰۰ می‌باشد. در نهایت، شکل (۲۰) نشان از این دارد که مدل حالت کاملا مرطوب را برای تمام نقاط کشور در دی ۱۴۰۰ پیش‌بینی کرده است. به مانند تفسیر شکل (۱۸) علت برقراری شرایط مرطوب‌تر برای استان‌های جنوبی کشور که غالبا آب و هوایی گرم و خشک دارند، شرایط بارش و دمایی بسیار متفاوت برای این استان‌ها در این بازه از سال نسبت به سایر اوقات سال می‌باشد که منجر به افزایش نسبی مقادیر شاخص SPEI برای این مناطق در این ماه شده است.



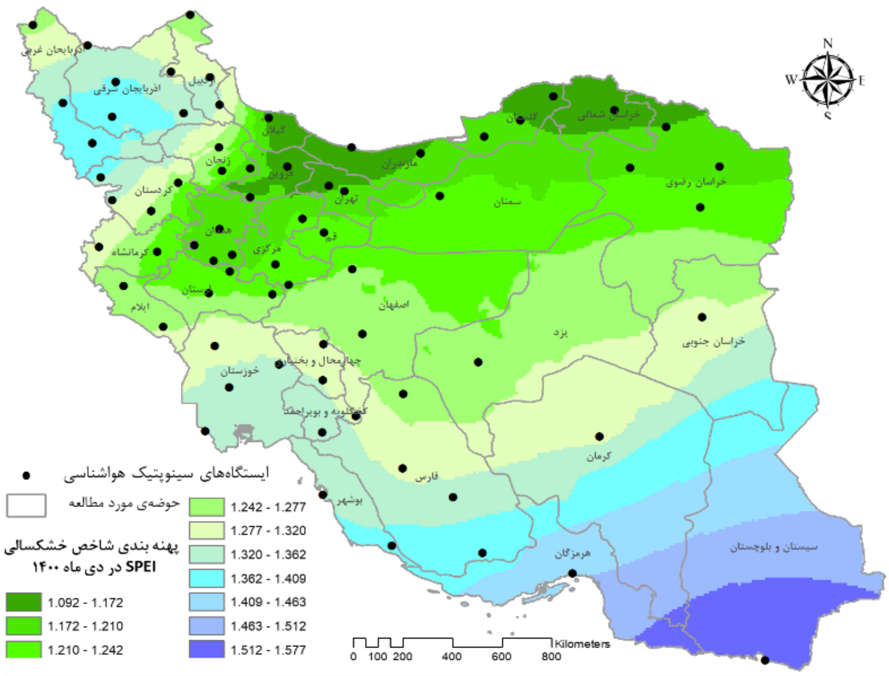
شکل ۱۷ نقشه پهنه‌بندی خشکسالی کشاورزی در فروردین ماه ۱۴۰۰



شکل ۱۸ نقشه پهنه‌بندی خشکسالی کشاورزی در تیر ماه ۱۴۰۰



شکل ۱۹ نقشه پهنه‌بندی خشکسالی کشاورزی در مهر ماه ۱۴۰۰



شکل ۲۰ نقشه پهنه‌بندی خشکسالی کشاورزی در دی ماه ۱۴۰۰

نتیجه‌گیری

در پژوهش حاضر سعی بر ایجاد راهکاری مناسب برای پیش‌بینی خشکسالی به عمل آمده است. در این مطالعه از چهار الگوریتم یادگیری ماشین به نام‌های جنگل تصادفی تصحیح بایاس شده، جنگل تصادفی، ماشین بردار پشتیبان و پرسپترون چندلایه به منظور ایجاد مدل پیش‌بینی خشکسالی استفاده گردید که نتایج آن به شرح زیر به دست آمد: ۱) برای ایستگاه‌های متعلق به خوشه ۱، که میانگین بارش سالانه‌ای در حدود ۹۵۰ میلی‌متر دارند، الگوریتم جنگل تصادفی تصحیح بایاس شده با دقت ۹۸/۵۸ درصد بهترین عملکرد را نشان داد. پس از

آن الگوریتم‌های جنگل تصادفی، پرسپترون چندلایه و ماشین بردار پشتیبان به ترتیب و با فاصله اندک نسبت به دقت الگوریتم جنگل تصادفی تصحیح بایاس شده، در رده‌های بعدی قرار گرفتند. ۲) برای ایستگاه‌های متعلق به خوشه ۲، که میانگین بارش سالانه آن‌ها کمتر از ۳۰۰ میلی‌متر می‌باشد، نیز الگوریتم جنگل تصادفی تصحیح بایاس شده با دقت ۹۸/۰۹ درصد بهترین عملکرد را نشان داد. پس از آن الگوریتم‌های ماشین بردار پشتیبان، جنگل تصادفی و پرسپترون چندلایه به ترتیب در رده‌های بعدی قرار گرفتند. ۳) بررسی‌ها نشان داد که مدل پیش‌بینی خشکسالی که با الگوریتم جنگل تصادفی تصحیح بایاس شده ایجاد شده است، قابلیت پیش‌بینی بسیار دقیقی هم در اقلیم پر باران و هم در اقلیم خشک و نیمه‌خشک دارد. ۴) از نتایج به دست آمده در این مدل‌سازی اقدام به تهیه نقشه‌های پهنه‌بندی خشکسالی شد. این نقشه‌ها که ابزاری مناسب برای شناسایی و طبقه‌بندی مناطق مستعد به شرایط خشکسالی می‌باشند، می‌توانند اطلاعات ارزشمندی را در رابطه با توزیع و شدت خشکسالی در مناطق مختلف در اختیار ذی‌نفعان، از جمله سیاست‌گذاران، کشاورزان و برنامه‌ریزان منابع آب قرار می‌دهد.

واژه‌نامه

شاخص شدت خشکسالی پالمر	PDSI
شاخص بارندگی استاندارد شده	SPI
شاخص بارش ـ تبخیر و تعرق استاندارد شده	SPEI
شاخص تفاوت پوشش گیاهی نرمال شده	NDVI
شاخص تفاوت خشکسالی نرمال شده	NDDI
شاخص خشکسالی نرمال شده چندباندی	NMDI
پرسپترون چندلایه	MLP
ماشین بردار پشتیبان	SVM
جنگل تصادفی	RF
جنگل تصادفی تصحیح بایاس شده	BRF
میانگین مربعات خطا	MSE
ریشه‌ی میانگین مربعات خطا	RMSE
میانگین خطای مطلق	MAE
ضریب تعیین	R

مراجع

[1] D. Dutta, A. Kundu, and N. R. Patel, "Predicting agricultural drought in eastern Rajasthan of India using NDVI and standardized precipitation index," *Geocarto International*, vol. 28, no. 3, pp. 192–209, 2013. https://doi.org/10.1080/10106049.2012.679975

[2] N. M. Alam *et al.*, "Evaluation of drought using SPEI drought class transitions and log-linear models for different agro-ecological regions of India," *Physics and Chemistry of the Earth, Parts A/B/C*, vol. 100, pp. 31–43, 2017. https://doi.org/10.1016/j.pce.2017.02.008

[3] W. C. Palmer, *Meteorological drought*, vol. 30. US Department of Commerce, Weather Bureau, 1965.

[4] T. B. McKee, N. J. Doesken, and J. Kleist, "The relationship of drought frequency and duration to time scales," in *Eighth Conference on Applied Climatology*, 1993, vol. 17, no. January, pp. 17–22.

[5] S. M. Vicente-Serrano, S. Beguería, and J. I. López-Moreno, "A Multiscalar Drought Index Sensitive to Global Warming: The Standardized Precipitation Evapotranspiration Index," *Journal of Climate*, vol. 23, no. 7, pp. 1696–1718, 2010. https://doi.org/10.1175/2009JCLI2909.1

[6] J. W. Rouse, R. H. Haas, J. A. Schell, and D. W. Deering, "Monitoring vegetation systems in the Great Plains with ERTS," *NASA Spec. Publ*, vol. 351, no. 1, p. 309, 1974.

[7] Y. Gu, J. F. Brown, J. P. Verdin, and B. Wardlow, "A five-year analysis of MODIS NDVI and NDWI for grassland drought assessment over the central Great Plains of the United States," *Geophysical research letters*, vol. 34, no. 6, 2007.

https://doi.org/10.1029/2006GL029127

[8] L. Wang and J. J. Qu, "NMDI : A normalized multi-band drought index for monitoring soil and vegetation moisture with satellite remote sensing," Geophysical research letters, vol. 34, no. July, pp. 1–5, 2007. https://doi.org/10.1029/2007GL031021

[9] J. Rhee, J. Im, and G. J. Carbone, "Monitoring agricultural drought for arid and humid regions using multi-sensor remote sensing data," *Remote Sensing of Environment*, vol. 114, no. 12, pp. 2875–2887, 2010. https://doi.org/10.1016/j.rse.2010.07.005

[10] Z. M. Yaseen, M. Ali, A. Sharafati, N. Al-Ansari, and S. Shahid, "Forecasting standardized precipitation index using data intelligence models: regional investigation of Bangladesh," *Scientific Reports*, vol. 11, no. 1, p. 3435, 2021. https://doi.org/10.1038/s41598-021-82977-9

[11] M. R. Kousari, M. E. Hosseini, H. Ahani, and H. Hakimelahi, "Introducing an operational method to forecast long-term regional drought based on the application of artificial intelligence capabilities," *Theoretical and Applied Climatology*, vol. 127, no. 1–2. pp. 361–380, 2017. https://doi.org/10.1007/s00704-015-1624-6

[12] M. Mokhtarzad, F. Eskandari, N. Jamshidi Vanjani, and A. Arabasadi, "Drought forecasting by ANN, ANFIS, and SVM and comparison of the models," *Environ. Earth Sci.*, vol. 76, no. 21, p. 729, 2017. https://doi.org/10.1007/s12665-017-7064-0

[13] A. Abbasi, K. Khalil, J. Behmanesh, and A. Shirzad, "Application of Support Vector Machine and Bayesian Network method in agricultural drought prediction," *Watershed Engineering and Management*, vol. 12, no. 1, pp. 107–124, 2020. https://doi.org/10.22092/ijwmse.2019.122619.1519

[14] F. Tufaner and A. Özbeyaz, "Estimation and easy calculation of the Palmer Drought Severity Index from the meteorological data by using the advanced machine learning algorithms," *Environmental Monitoring and Assessment*, vol. 192, no. 9, 2020. https://doi.org/10.1007/s10661-020-08539-0

[15] P. Feng, B. Wang, D. L. Liu, and Q. Yu, "Machine learning-based integration of remotely-sensed drought factors can improve the estimation of agricultural drought in South-Eastern Australia," *Agricultural Systems*, vol. 173. pp. 303–316, 2019. https://doi.org/10.1016/j.agsy.2019.03.015

[16] M. Zoljoodi and A. Didevarasl, "Evaluation of Spatial-Temporal Variability of Drought Events in Iran Using Palmer Drought Severity Index and Its Principal Factors (through 1951-2005)," *Atmospheric and Climate Sciences*, vol. 03, no. 02, pp. 193–207, 2013. http://dx.doi.org/10.4236/acs.2013.32021

[17] A. Darvishi, S. Arkhi, and A. Ebrahimi, "Risk and disaster management to mitigate the effects of droughts in Iran," in *Proceeding of the Conference on Drought in Charmahal-Bakhtiari, Shahrekord, The University of Shahrekord*, 2008, pp. 22–23.

[18] Iran's Statistical Centre, "Shares and rankings of provinces based on important variables in the agricultural sector," 2023.

[19] J. A. Hartigan and M. A. Wong, "Algorithm AS 136: A K-Means Clustering Algorithm," *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, vol. 28, no. 1, pp. 100–108, 1979. https://doi.org/10.2307/2346830

[20] C. W. Thornthwaite, "An Approach toward a Rational Classification of Climate," *Geogr. Rev.*, vol. 38, no. 1, pp. 55–94, 1948. https://doi.org/10.2307/210739

[21] R. Tibshirani, "Regression Shrinkage and Selection via the Lasso," *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 58, no. 1, pp. 267–288, 1996. https://doi.org/10.1111/j.2517-6161.1996.tb02080.x

[22] F. Rosenblatt, *Principles of neurodynamics: Perceptrons and the theory of brain mechanisms*, vol. 55. Spartan books Washington, DC, 1962.

[23] A. Aghasi, M. Torbatinejad, and A. Bagherzadeh, "Modelling of rutting behavior of modified polymer asphalt mixture using artificial neural network," *Ferdowsi Civil Engineering*, vol. 35, no. 3, pp. 69–88, 2022. https://doi.org/10.22067/jfcei.2022.76444.1135

[24] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, pp. 273–297, 1995. https://doi.org/10.1007/BF00994018

[25] L. Breiman, "Random Forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001. https://doi.org/10.1023/A:1010933404324

[26] J. Song, "Bias corrections for Random Forest in regression using residual rotation," *Journal of the Korean Statistical Society*, vol. 44, no. 2, pp. 321–326, 2015. https://doi.org/10.1016/j.jkss.2015.01.003

[27] G. Zhang and Y. Lu, "Bias-corrected random forests in regression," *Journal of Applied Statistics*, vol. 39, no. 1, pp. 151–160, 2012. https://doi.org/10.1080/02664763.2011.578621

[28] Iran Meteorological Organization, "Weather data acquisition system," *Ferdowsi Civil Engineering*, 2024. https://doi.org/10.22067/jfcei.2024.88855.1311

[29] G. J. Huffman E.F. Stocker, D.T. Bolvin, E.J. Nelkin, Jackson Tan, "GPM IMERG Final Precipitation L3 Half Hourly 0.1 degree x 0.1 degree V06." *NASA Goddard Earth Sciences Data and Information Services Cente*, 2019.

[30] Z. Wan, S. Hook, and G. Hulley, "MODIS/Terra Land Surface Temperature/Emissivity Daily L3 Global 1km SIN Grid V061." N*ASA EOSDIS Land Processes Distributed Active Archive Center*, 2021.

[31] E. Vermote and R. Wolfe, "MODIS/Terra Surface Reflectance Daily L2G Global 1km and 500m SIN Grid V061." *NASA EOSDIS Land Processes Distributed Active Archive Center*, 2021.

[32] Z. Zhao, S. Xu, B. H. Kang, M. M. J. Kabir, Y. Liu, and R. Wasinger, "Investigation and improvement of multi-layer perception neural networks for credit scoring," *Expert Systems with Applications*, vol. 42, no. 7, pp. 3508–3516, 2015. https://doi.org/10.1016/j.eswa.2014.12.006

[33] A. Rani, N. Kumar, J. Kumar, J. Kumar, and N. K. Sinha, "Chapter 6 - Machine learning for soil moisture assessment," *In Deep learning for sustainable agriculture*, pp. 143-168, 2022. https://doi.org/10.1016/B978-0-323-85214-2.00001-X

[34] A. Afzal *et al.*, "Response surface analysis, clustering, and random forest regression of pressure in suddenly expanded high-speed aerodynamic flows," *Aerospace Science and Technology*, vol. 107, p. 106318, 2020. https://doi.org/10.1016/j.ast.2020.106318